OM protein - protein search, using sw model

Run on: May 18, 2006, 15:32:22 ; Search time 194 Seconds
(without alignments)
539.704 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_8:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*
9: geneseqp2005s:*
10: geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1198	100.0	229	4	AAM40223	Aam40223 Human pol
2	1198	100.0	229	5	ABG34856	Abg34856 Human can
3	1198	100.0	231	4	AAM42009	Aam42009 Human pol
4	138.5	11.6	164	5	ABG34855	Abg34855 Human cDN
5	113	9.4	233	4	AAM93668	Aam93668 Human pol
6	113	9.4	233	4	AAG93290	Aag93290 Human pro
7	113	9.4	233	7	ADJ69163	Adj69163 Human hea
8	113	9.4	233	8	ADL31517	Adl31517 Human pro
9	92	7.7	695	5	ABB54167	Abb54167 Lactococc
10	92	7.7	695	8	ADS29356	Ads29356 Bacterial
11	90	7.5	269	4	AAO08972	Aao08972 Human pol
12	89	7.4	572	8	ABM83354	Abm83354 Human dia
13	89	7.4	623	4	AAB93182	Aab93182 Human pro
14	89	7.4	623	5	ABB97233	Abb97233 Novel hum
15	89	7.4	652	4	AAB93168	Aab93168 Human pro
16	89	7.4	664	4	AAB83843	Aab83843 Amino aci
17	89	7.4	664	4	AAB20219	Aab20219 Human Chf
18	88.5	7.4	388	8	ADQ08702	Adq08702 Ciona int
19	87.5	7.3	1245	7	ADC32818	Adc32818 Human nov
20	86.5	7.2	470	8	ABM83305	Abm83305 Human dia
21	86.5	7.2	669	6	ABU23878	Abu23878 Protein e
22	86	7.2	158	7	ADH88711	Adh88711 Enterococ
23	85.5	7.1	571	8	ADU07910	Adu07910 Amino aci
24	84.5	7.1	485	5	ABP51336	Abp51336 Human MDD
25	84	7.0	496	4	AAB95788	Aab95788 Human pro
26	84	7.0	496	7	ADC31196	Adc31196 Human nov
27	83.5	7.0	479	4	AAU39678	Aau39678 Propionib
28	83.5	7.0	479	6	ABM36197	Abm36197 Propionib
29	83	6.9	598	2	AAW59461	Aaw59461 Microbial
30	83	6.9	598	2	AAW59463	Aaw59463 Microbial
31	83	6.9	598	2	AAW59457	Aaw59457 Microbial
32	83	6.9	598	2	AAW59458	Aaw59458 Microbial
33	83	6.9	598	2	AAW59459	Aaw59459 Microbial
34	83	6.9	598	2	AAW59462	Aaw59462 Microbial
35	83	6.9	598	2	AAW59460	Aaw59460 Microbial
36	83	6.9	598	2	AAW59464	Aaw59464 Microbial
37	83	6.9	599	2	AAW59496	Aaw59496 Microbial
38	83	6.9	606	1	AAP90110	Aap90110 Polypepti
39	83	6.9	607	2	AAW59456	Aaw59456 Microbial
40	83	6.9	607	2	AAW59465	Aaw59465 Microbial
41	83	6.9	663	6	ABU23749	Abu23749 Protein e
42	82.5	6.9	278	8	ADX72944	Adx72944 Plant ful
43	82.5	6.9	325	3	AAY90242	Aay90242 Human cyc
44	82.5	6.9	493	8	ADN72405	Adn72405 Thale cre
45	82.5	6.9	493	9	AEA16977	Aea16977 Arabidops

ALIGNMENTS

RESULT 1
AAM40223
ID AAM40223 standard; protein; 229 AA.
XX
AC AAM40223;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 3368.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI59379.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.

XX
PS Example 5; SEQ ID NO 3368; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 229 AA;

Query Match 100.0%; Score 1198; DB 4; Length 229;
Best Local Similarity 100.0%; Pred. No. 1e-126;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

RESULT 2
ABG34856
ID ABG34856 standard; protein; 229 AA.
XX
AC ABG34856;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human cancer related protein encoded by cDNA 85P1B3.
XX
KW Human; cytostatic; 85P1B3; cancer; immunogen; chromosome 15q14.
XX
OS Homo sapiens.
XX
PN WO200218578-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001WO-US026838.
XX
PR 28-AUG-2000; 2000US-0228432P.
XX
PA (AGEN-) AGENSYS INC.
XX
PI Raitano AB, Faris M, Hubert RS, Afar D, Ge W, Challita-Eid P;
PI Jakobovits A;
XX
DR WPI; 2002-382963/41.
DR N-PSDB; ABK70506.
XX
PT Composition for modulating the status of 85P1B3 protein or a molecule
PT comprising a substance e.g. antibody specific to, nucleic acid encoding,
PT or ribozyme of 85P1B3.
XX
PS Claim 34; Fig 2; 201pp; English.
XX
CC The invention relates to a composition comprising a substance that
CC modulate the status of 85P1B3, where the status of a cell expresses
CC 85P1B3 gene product is modulated. Also included are a composition
CC comprising a peptide region of 5 amino acids of the 85P1B3 protein, in
CC any whole number increment up to 229 that includes an aa position
CC selected from an aa position having a value greater than 0.5 in the
CC Hydrophilicity profile, an aa position having a value less than 0.5 in
CC the hydropathicity profile, an aa position having a value greater than
CC 0.5 in the percent accessible residue profile, an aa position having a
CC value greater than 0.5 in the average flexibility profile, or an aa
CC position having a value greater than 0.5 in the beta-turn profile; a
CC polynucleotide that encodes analogue peptide of 8, 9, 10 or 11 contiguous
CC residues of the 85P1B3 protein; a recombinant protein comprising the
CC antigen-binding region of a monoclonal antibody; a non-human transgenic
CC animal that produces an antibody that binds to the 85P1B3 protein; a
CC hybridoma that produces antibody specific to the protein; a single chain
CC monoclonal antibody (MAb) that comprises the variable domains of the
CC heavy and monoclonal antibodies specific to the protein; a vector
CC comprising a polynucleotide that encodes the MAb; inhibiting growth of
CC cancer cells or treating a patient who bears cancer cells that expresses
CC the protein, by administering the protein, antibody, polynucleotide
CC encoding the protein, antisense polynucleotide to the polynucleotide,
CC ribozyme that cleaves the polynucleotide and T cells that specifically
CC recognize the protein; and generating a mammalian immune response
CC directed to the protein exposing cells of the mammal's immune system to
CC an immunogenic portion of the protein or polynucleotide. The composition,
CC which comprises an antibody specific to the protein, is useful for
CC delivering a cytotoxic agent to a cell that expresses the protein by
CC providing a cytotoxic agent conjugated to antibody and exposing the cell
CC to the antibody-agent conjugate. The methods are useful for inhibiting
CC growth of cancer cells or treating a patient who bears cancer cells that
CC expresses the protein, for generating a mammalian immune response
CC directed to the protein, for detecting the presence of the protein or
CC polynucleotide in a biological sample in a patient who has or who is
CC suspected of having cancer and for monitoring 85P1B3 in a biological
CC sample from a patient who has or who is suspected of having cancer. The
CC gene for 85P1B3 is located on human chromosome 15q14. The present
CC sequence is the 85P1B3 protein
XX
SQ Sequence 229 AA;

Query Match 100.0%; Score 1198; DB 5; Length 229;
Best Local Similarity 100.0%; Pred. No. 1e-126;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

RESULT 3
AAM42009
ID AAM42009 standard; protein; 231 AA.
XX
AC AAM42009;
XX

DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6940.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI61165.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 2; SEQ ID NO 6940; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 231 AA;

Query Match 100.0%; Score 1198; DB 4; Length 231;
Best Local Similarity 100.0%; Pred. No. 1e-126;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
Db 3 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 62

Qy 61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
Db 63 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 122

Qy 121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
Db 123 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 182

Qy 181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
Db 183 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 231

RESULT 4
ABG34855
ID ABG34855 standard; protein; 164 AA.
XX
AC ABG34855;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human cDNA 85P1B3 splice variant, open reading frame #3.
XX
KW Human; cytostatic; 85P1B3; cancer; immunogen; chromosome 15q14.
XX
OS Homo sapiens.
XX
PN WO200218578-A2.
XX
PD 07-MAR-2002.
XX
PF 28-AUG-2001; 2001WO-US026838.
XX
PR 28-AUG-2000; 2000US-0228432P.
XX
PA (AGEN-) AGENSYS INC.
XX
PI Raitano AB, Faris M, Hubert RS, Afar D, Ge W, Challita-Eid P;
PI Jakobovits A;
XX
DR WPI; 2002-382963/41.
DR N-PSDB; ABK70504.
XX
PT Composition for modulating the status of 85P1B3 protein or a molecule
PT comprising a substance e.g. antibody specific to, nucleic acid encoding,
PT or ribozyme of 85P1B3.
XX
PS Example 38; Page 124; 201pp; English.
XX
CC The invention relates to a composition comprising a substance that
CC modulate the status of 85P1B3, where the status of a cell expresses
CC 85P1B3 gene product is modulated. Also included are a composition
CC comprising a peptide region of 5 amino acids of the 85P1B3 protein, in
CC any whole number increment up to 229 that includes an aa position
CC selected from an aa position having a value greater than 0.5 in the
CC Hydrophilicity profile, an aa position having a value less than 0.5 in
CC the hydropathicity profile, an aa position having a value greater than
CC 0.5 in the percent accessible residue profile, an aa position having a
CC value greater than 0.5 in the average flexibility profile, or an aa
CC position having a value greater than 0.5 in the beta-turn profile; a
CC polynucleotide that encodes analogue peptide of 8, 9, 10 or 11 contiguous
CC residues of the 85P1B3 protein; a recombinant protein comprising the
CC antigen-binding region of a monoclonal antibody; a non-human transgenic
CC animal that produces an antibody that binds to the 85P1B3 protein; a
CC hybridoma that produces antibody specific to the protein; a single chain
CC monoclonal antibody (MAb) that comprises the variable domains of the
CC heavy and monoclonal antibodies specific to the protein; a vector
CC comprising a polynucleotide that encodes the MAb; inhibiting growth of
CC cancer cells or treating a patient who bears cancer cells that expresses
CC the protein, by administering the protein, antibody, polynucleotide
CC encoding the protein, antisense polynucleotide to the polynucleotide,
CC ribozyme that cleaves the polynucleotide and T cells that specifically
CC recognize the protein; and generating a mammalian immune response
CC directed to the protein exposing cells of the mammal's immune system to
CC an immunogenic portion of the protein or polynucleotide. The composition,
CC which comprises an antibody specific to the protein, is useful for
CC delivering a cytotoxic agent to a cell that expresses the protein by

```
CC   providing a cytotoxic agent conjugated to antibody and exposing the cell
CC   to the antibody-agent conjugate. The methods are useful for inhibiting
CC   growth of cancer cells or treating a patient who bears cancer cells that
CC   expresses the protein, for generating a mammalian immune response
CC   directed to the protein, for detecting the presence of the protein or
CC   polynucleotide in a biological sample in a patient who has or who is
CC   suspected of having cancer and for monitoring 85P1B3 in a biological
CC   sample from a patient who has or who is suspected of having cancer. The
CC   gene for 85P1B3 is located on human chromosome 15q14. The present
CC   sequence is the 85P1B3 protein
XX
SQ   Sequence 164 AA;

  Query Match             11.6%;  Score 138.5;  DB 5;  Length 164;
  Best Local Similarity   82.1%;  Pred. No. 9e-07;
  Matches   32;  Conservative    1;  Mismatches    1;  Indels    5;  Gaps    1;

Qy        162 FCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              | |||     |||||||||||||||||||||||||||:|
Db          2 FFLSS-----YLLKTKAIVNASEMDIQNVPLSEKIAEVK 35


RESULT 5
AAM93668
ID   AAM93668 standard; protein; 233 AA.
XX
AC   AAM93668;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polypeptide, SEQ ID NO: 3550.
XX
KW   Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS   Homo sapiens.
XX
PN   EP1130094-A2.
XX
PD   05-SEP-2001.
XX
PF   07-JUL-2000; 2000EP-00114089.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183765.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2001-524255/58.
DR   N-PSDB; AAK94603.
XX
PT   830 Primers useful for synthesizing full length cDNA clones and their use
PT   in genetic manipulation.
XX
PS   Claim 8; SEQ ID NO 3550; 1380pp + Sequence Listing; English.
XX
CC   The invention relates to primers for synthesising full length cDNA
CC   clones. 830 cDNA molecules encoding a human protein have been isolated
CC   and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have
CC   been determined. Primers for synthesising the full length cDNA are useful
CC   for clarifying the function of the protein encoded by the cDNA. The full
CC   length clones were obtained by construction of full length enriched cDNA
CC   libraries that were synthesised by the oligo-capping method. The primers
CC   enable the production of the full length cDNA easily without any special
CC   methods. The present sequence is a polypeptide encoded by a full length
CC   human cDNA of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in CD-ROM
CC   format directly from EPO
XX
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 4;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0012;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  | |    : :  | |  || ||  :        ||  || |: |    | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :        |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 6
AAG93290
ID   AAG93290 standard; protein; 233 AA.
XX
AC   AAG93290;
XX
DT   13-SEP-2001  (first entry)
XX
DE   Human protein HP10650.
XX
KW   Human; gene therapy; tumour.
XX
OS   Homo sapiens.
XX
PN   WO200142302-A1.
XX
PD   14-JUN-2001.
XX
PF   06-DEC-2000; 2000WO-JP008631.
XX
PR   06-DEC-1999;   99JP-00346863.
PR   06-DEC-1999;   99JP-00346864.
PR   08-FEB-2000; 2000JP-00031062.
PR   10-FEB-2000; 2000JP-00034090.
PR   10-FEB-2000; 2000JP-00034091.
PR   14-FEB-2000; 2000JP-00035829.
PR   14-FEB-2000; 2000JP-00035899.
PR   14-MAR-2000; 2000JP-00071161.
PR   30-MAY-2000; 2000JP-00160851.
XX
PA   (NISC-) JAPAN SCI & TECHNOLOGY CORP.
XX
PI   Kato S,  Eguchi C,  Saeki M;
XX
DR   WPI; 2001-381646/40.
DR   N-PSDB; AAH68575.
XX
PT   Human protein originated from tumor cell line, applicable as drug,
PT   reagent for studying intracellular protein networks and protein source
PT   for drug screening, also encoded cDNA for gene diagnosis and gene
PT   therapy.
XX
PS   Claim 1; Page 303-304; 471pp; Japanese.
XX
CC   The present sequence is a human protein. The human protein, preferably
CC   originated from tumour cell line, is applicable as a drug, a reagent for
CC   studying intracellular protein networks and a protein source for
CC   screening proteins for binding low molecular weight drugs. The human
CC   protein coding sequence is useful for gene diagnosis and gene therapy,
CC   expression vectors and transformant cells for detection of ligands and
CC   receptors
```

```
XX
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 4;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0012;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||. :||  |  |    : :  | |  || ||  :      ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 7
ADJ69163
ID   ADJ69163 standard; protein; 233 AA.
XX
AC   ADJ69163;
XX
DT   06-MAY-2004  (first entry)
XX
DE   Human heat mitochondrial protein as a therapeutic target SeqID969.
XX
KW   mitochondrial; human; screening assay; diabetes mellitus;
KW   Huntington's disease; osteoarthritis;
KW   Leber's hereditary optic neuropathy; LHON;
KW   mitochondrial encephalopathy lactic acidosis and stroke; MELAS;
KW   myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;
KW   neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;
KW   osteopathic; ophthalmological; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2003087768-A2.
XX
PD   23-OCT-2003.
XX
PF   04-APR-2003; 2003WO-US010870.
XX
PR   12-APR-2002; 2002US-0372843P.
PR   17-JUN-2002; 2002US-0389987P.
PR   20-SEP-2002; 2002US-0412418P.
XX
PA   (MITO-) MITOKOR.
PA   (BUCK-) BUCK INST AGE RES.
XX
PI   Ghosh SS,  Fahy ED,  Zhang B,  Gibson BW,  Taylor SW,  Glenn GM;
PI   Warnock DE;
XX
DR   WPI; 2003-845369/78.
XX
PT   Identifying a mitochondrial target for drug screening assays and for
PT   treating diseases associated with altered mitochondrial function,
PT   comprises detecting a modified polypeptide in a sample and correlating
PT   with the disease.
XX
PS   Claim 1; SEQ ID NO 969; 180pp; English.
XX
CC   This invention relates to novel mitochondrial targets that can be used
CC   for therapeutic intervention in treating a disease associated with
CC   altered mitochondrial function. Specifically, it refers to a method for
CC   identifying proteins of the human heart mitochondrial proteome that are
CC   useful for drug screening assays, as well as therapeutic targets. The
CC   present invention describes a method for identifying such proteins that
CC   can be used in the treatment of various diseases associated with altered
CC   mitochondrial function including diabetes mellitus, Huntington's disease,
CC   osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial
CC   encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy
CC   ragged red fibre syndrome (MERRF) or cancer. Accordingly, these
CC   compositions have neuroprotective, nootropic, antidiabetic,
CC   anticonvulsant, antiarthritic, osteopathic, ophthalmological and
CC   cytostatic activities. This polypeptide sequence is a human heart
CC   mitochondrial protein of the invention.
XX
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 7;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0012;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :      ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 8
ADL31517
ID   ADL31517 standard; protein; 233 AA.
XX
AC   ADL31517;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Human protein encoded by a full length cDNA clone SeqID 3550.
XX
KW   human; medicine; signal transduction; glycoprotein; transcription;
KW   oligo-capping method.
XX
OS   Homo sapiens.
XX
PN   EP1396543-A2.
XX
PD   10-MAR-2004.
XX
PF   07-JUL-2000; 2003EP-00025638.
XX
PR   08-JUL-1999;   99JP-00194486.
PR   11-JAN-2000; 2000JP-00118774.
PR   02-MAY-2000; 2000JP-00183865.
PR   07-JUL-2000; 2000EP-00114089.
XX
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Ota T,  Nishikawa T,  Isogai T,  Hayashi K,  Ishii S,  Kawai Y;
PI   Wakamatsu A,  Sugiyama T,  Nagai K,  Kojima S,  Otsuki T,  Koga H;
XX
DR   WPI; 2004-204755/20.
DR   N-PSDB; ADL31516.
XX
PT   New oligonucleotide primers (830 cDNAs) useful for synthesizing full
PT   length human cDNAs.
XX
PS   Example 1; SEQ ID NO 3550; 1340pp; English.
```

```
XX
CC   This invention relates to a novel primers useful for synthesising full
CC   length cDNA molecules that encode human proteins. Specifically, it refers
CC   to secretory or membrane proteins that are potential therapeutic agents/
CC   target molecules in the field of medicine, and in particular genes
CC   encoding proteins that are associated with signal transduction,
CC   glycoproteins and transcription. The present invention describes a method
CC   for efficiently cloning a full length human cDNA from both the 5' and 3'
CC   ends using the oligo-capping method. This polypeptide sequence is a full
CC   length human protein of the invention.
XX
SQ   Sequence 233 AA;

  Query Match             9.4%;  Score 113;  DB 8;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0012;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVPLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :        |      | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
              |    |   | |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 9
ABB54167
ID   ABB54167 standard; protein; 695 AA.
XX
AC   ABB54167;
XX
DT   29-AUG-2003  (revised)
DT   16-MAY-2002  (first entry)
XX
DE   Lactococcus lactis protein yihC.
XX
KW   Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS   Lactococcus lactis; IL1403.
XX
PN   FR2807446-A1.
XX
PD   12-OCT-2001.
XX
PF   11-APR-2000; 2000FR-00004630.
XX
PR   11-APR-2000; 2000FR-00004630.
XX
PA   (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
PI   Bolotine A,  Sorokine A,  Renault P,  Ehrlich SD;
XX
DR   WPI; 2002-043418/06.
XX
PT   New nucleotide sequence useful in the identification or Lactococcus
PT   lactis and related species.
XX
PS   Claim 6; SEQ ID NO 869; 2504pp; French.
XX
CC   The present invention is related to a Lactococcus lactis nucleotide
CC   sequence (ABA90521) and related proteins (ABB53300-ABB55621). The nucleic
CC   acid sequence is useful in the detection and/or amplification of nucleic
CC   acid sequence, particularly to identify Lactococcus lactis or related
CC   species. The proteins of the invention are useful for the biosynthesis or
CC   biodegradation of a composition of interest. The invention helps research
CC   in lactic bacteria, particularly useful in the production of yogurt and
CC   cheese. Note: The sequence data for this patent is based on equivalent
CC   patent WO200177334 (published 18-OCT-2001) which is available in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences. (Updated on 29-AUG-2003 to
CC   standardise OS field)
XX
SQ   Sequence 695 AA;

  Query Match             7.7%;  Score 92;  DB 5;  Length 695;
  Best Local Similarity   25.3%;  Pred. No. 1.5;
  Matches   38;  Conservative   28;  Mismatches   50;  Indels   34;  Gaps    7;

Qy        101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
              || ::|              ||  | :| ||      ::   : || | :  || :| : :| |
Db        101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy        161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI-------------QNVPLSE-K 195
              |           |    :: :|:: |||: | :   |:            ::| ||| |
Db        154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy        196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
              : :|   :|  : :: :   ||||   |:|
Db        214 VGDL--LLVKENEKIPADGLILSEALVDES 241


RESULT 10
ADS29356
ID   ADS29356 standard; protein; 695 AA.
XX
AC   ADS29356;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #18389.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; plant growth regulator;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
PA   (CAOY/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y,  Hinkle GJ,  Slater SC,  Chen X,  Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 18389; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
```

CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 695 AA;

Query Match 7.7%; Score 92; DB 8; Length 695;
Best Local Similarity 25.3%; Pred. No. 1.5;
Matches 38; Conservative 28; Mismatches 50; Indels 34; Gaps 7;

```
Qy        101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
              || ::|          ||  | :| ||     ::   : || | :  || :| : :| |
Db        101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy        161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI-------------QNVPLSE-K 195
              |          |    :: :|:: |||: | :   |:             ::| ||| |
Db        154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy        196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
              : :|   :|  : :: :   ||||   |:|
Db        214 VGDL--LLVKENEKIPADGLILSEALVDES 241
```

RESULT 11
AAO08972
ID AAO08972 standard; protein; 269 AA.
XX
AC AAO08972;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 22864.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US004927.
XX
PR 28-FEB-2000; 2000US-00515126.
PR 18-MAY-2000; 2000US-00577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
DR N-PSDB; AAI88903.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing diagnosing
PT and treating e.g. leukemia, inflammation and immune disorders.
XX
PS Claim 20; SEQ ID NO 22864; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation. Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 269 AA;

Query Match 7.5%; Score 90; DB 4; Length 269;
Best Local Similarity 23.0%; Pred. No. 0.61;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

```
Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|:  |     | ||:::    ::||  |
Db         39 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 95

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||          |||  |   |    || |  :|
Db         96 PMPDRRAEREQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 139

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|               | :
Db        140 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 183

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
              :|  || | ||      | ::||    ::    :    ||:| ||
Db        184 LMALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 229
```

RESULT 12
ABM83354
ID ABM83354 standard; protein; 572 AA.
XX
AC ABM83354;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic pprotein SEQ ID NO:3603.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; dithp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
PR 12-SEP-2002; 2002US-0410260P.
XX
PA (INCY-) INCYTE CORP.
XX
PI Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F;
PI Harthshorne TA, Suchorolski MT, Altus CM, Pitts SJ, Elder LV;
PI Mooney EM, Delegeane AM, Panesar IS, Banville SC, Reddy TP;
PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerstin EH;

PI Peralta CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LL;
PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton ES;
PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D;
PI Patury S, Shi X, Suarez CJ;
XX
DR WPI; 2004-329368/30.
DR N-PSDB; ACN42006.
XX
PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
PT in diagnosing a condition, disease or disorder associated with human
PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
PT in gene mapping.
XX
PS Claim 27; Page; 190pp; English.
XX
CC The invention relates to novel diagnostic and therapeutic polynucleotides
CC selected from one of the 2722 sequences defined in the specification. A
CC polynucleotide of the invention may have a use in gene therapy. The human
CC diagnostic and therapeutic polynucleotides (dithp) or polypeptides may be
CC used to diagnose a particular condition, disease or disorder associated
CC with human molecules, e.g. cell proliferative disorders,
CC autoimmune/inflammatory disorder, developmental disorder, endocrine
CC disorder, neurological disorders, gastrointestinal disorders, or
CC infections caused by virus, bacteria, fungi or parasite. The dithp
CC molecules may also be used in genetic mapping, in identifying individuals
CC from minute biological samples, in detecting single nucleotide
CC polymorphisms, as molecular weight markers, and for somatic or germline
CC gene therapy. The present sequence represents a dithp protein of the
CC invention. Note: The sequence data for this patent is not represented in
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at www.wipo.int/pct/en/sequences/listing.htm
XX
SQ Sequence 572 AA;

Query Match 7.4%; Score 89; DB 8; Length 572;
Best Local Similarity 23.0%; Pred. No. 2.5;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

```
Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|:  |     | ||:::     ::||  |
Db        342 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 398

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                  ::|   ||              |||  |    |    || |  :|
Db        399 PMPDRRAEREQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 442

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|                | :
Db        443 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 486

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
               :|  || | ||      | ::||    ::    :    ||:| ||
Db        487 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 532
```

RESULT 13
AAB93182
ID AAB93182 standard; protein; 623 AA.
XX
AC AAB93182;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12128.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 12128; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesising 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 623 AA;

Query Match 7.4%; Score 89; DB 4; Length 623;
Best Local Similarity 23.0%; Pred. No. 2.9;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

```
Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|:  |     | ||:::     ::||  |
Db        393 YRRQAAQPPH---CPAPEGEPGAPQALGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQ 449

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                  ::|   ||              |||  |    |    || |  :|
Db        450 PMPDRRAEREQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 493

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|                | :
Db        494 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 537

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
               :|  || | ||      | ::||    ::    :    ||:| ||
Db        538 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYQQNIPASE 583
```

RESULT 14
ABB97233
ID ABB97233 standard; protein; 623 AA.
XX
AC ABB97233;

```
XX
DT   27-JUN-2002  (first entry)
XX
DE   Novel human protein SEQ ID NO: 501.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US026015.
XX
PR   11-SEP-2000; 2000US-00659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32419.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis.
XX
PS   Example 2; SEQ ID NO 501; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC   and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC   rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention
XX
SQ   Sequence 623 AA;

  Query Match             7.4%;  Score 89;  DB 5;  Length 623;
  Best Local Similarity   23.0%;  Pred. No. 2.9;
  Matches   52;  Conservative   26;  Mismatches   74;  Indels   74;  Gaps   11;

Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|:  |     | ||:::     ::||  |
Db        393 YRRQAAQPPH---CPAPEGEPGAPQALGDAPPTSVSLTTAVQDYVCPLQGSHALCTCCFQ 449

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||              ||| |   |    || |  :|
Db        450 PMPDRRAEREQDPRVAPQ-------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 493

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|                | :
Db        494 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 537

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
               :|  || | ||      | ::||    ::    :    ||:| ||
Db        538 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYQQNIPASE 583


RESULT 15
AAB93168
ID   AAB93168 standard; protein; 652 AA.
XX
AC   AAB93168;
XX
DT   26-JUN-2001  (first entry)
XX
DE   Human protein sequence SEQ ID NO:12100.
XX
KW   Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS   Homo sapiens.
XX
PN   EP1074617-A2.
XX
PD   07-FEB-2001.
XX
PF   28-JUL-2000; 2000EP-00116126.
XX
PR   29-JUL-1999;   99JP-00248036.
PR   27-AUG-1999;   99JP-00300253.
PR   11-JAN-2000; 2000JP-00118776.
PR   02-MAY-2000; 2000JP-00183767.
PR   09-JUN-2000; 2000JP-00241899.
XX
PA   (HELI-) HELIX RES INST.
XX
PI   Ota T,  Isogai T,  Nishikawa T,  Hayashi K,  Saito K,  Yamamoto J;
PI   Ishii S,  Sugiyama T,  Wakamatsu A,  Nagai K,  Otsuki T;
XX
DR   WPI; 2001-318749/34.
XX
PT   Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT   length cDNAs defined in the specification, and for the detection and/or
PT   diagnosis of the abnormality of the proteins encoded by the full-length
PT   cDNAs.
XX
PS   Claim 8; SEQ ID NO 12100; 2537pp + Sequence Listing; English.
XX
CC   The present invention describes primer sets for synthesising 5602 full-
CC   length cDNAs defined in the specification. Where a primer set comprises:
CC   (a) an oligo-dT primer and an oligonucleotide complementary to the
CC   complementary strand of a polynucleotide which comprises one of the 5602
CC   nucleotide sequences defined in the specification, where the
CC   oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC   of an oligonucleotide comprising a sequence complementary to the
CC   complementary strand of a polynucleotide which comprises a 5'-end
CC   sequence and an oligonucleotide comprising a sequence complementary to a
CC   polynucleotide which comprises a 3'-end sequence, where the
CC   oligonucleotide comprises at least 15 nucleotides and the combination of
CC   the 5'-end sequence/3'-end sequence is selected from those defined in the
CC   specification. The primer sets can be used in antisense therapy and in
CC   gene therapy. The primers are useful for synthesising polynucleotides,
CC   particularly full-length cDNAs. The primers are also useful for the
CC   detection and/or diagnosis of the abnormality of the proteins encoded by
CC   the full-length cDNAs. The primers allow obtaining of the full-length
CC   cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC   AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC   represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC   oligonucleotides, all of which are used in the exemplification of the
CC   present invention
XX
SQ   Sequence 652 AA;

  Query Match             7.4%;  Score 89;  DB 4;  Length 652;
  Best Local Similarity   23.0%;  Pred. No. 3.1;
  Matches   52;  Conservative   26;  Mismatches   74;  Indels   74;  Gaps   11;

Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|:  |     | ||:::     ::||  |
Db        422 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 478

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||              ||| |   |    || |  :|
Db        479 PMPDRRAEREQDPRVAPQ-------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 522

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|                | :
```

```
Db        523 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML---------------TES 566

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
              :|  || | ||      | ::||    ::    :    ||:| ||
Db        567 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 612
```

Search completed: May 18, 2006, 15:35:47
Job time : 197 secs

GenCore version 5.1.8

OM protein - protein search, using sw model

Run on: May 18, 2006, 15:36:06 ; Search time 39 Seconds
(without alignments)
564.965 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	92	7.7	695	2	C86731	copper-potassium t
2	90.5	7.6	597	2	D71293	probable phosphori
3	89	7.4	306	2	T46399	hypothetical prote
4	88	7.3	250	2	A43623	kanamycin kinase (
5	86.5	7.2	669	2	A97229	NAD-dependent DNA
6	85	7.1	2194	1	JQ1977	glutamate synthase
7	84.5	7.1	361	2	AE2074	ferrichrome bindin
8	83	6.9	663	2	D97047	DNA ligase (NAD de
9	82.5	6.9	143	2	B69099	formate hydrogenly
10	82.5	6.9	493	2	A85433	sugar transporter
11	82	6.8	967	2	F87678	DNA polymerase I [
12	81.5	6.8	329	2	D90404	transport protein,
13	81.5	6.8	457	2	H85767	glucuronide permea
14	81.5	6.8	457	2	C90919	glucuronide permea
15	81.5	6.8	457	2	B64918	glucuronide permea
16	81.5	6.8	640	2	T41977	hypothetical prote
17	81.5	6.8	662	2	T44221	probable capsid pr
18	81	6.8	1418	2	S40764	hypothetical prote
19	80.5	6.7	766	2	B85440	receptor kinase-li
20	80.5	6.7	1711	1	A55148	protein-tyrosine-p
21	80	6.7	469	2	S55167	IME2-dependent sig
22	79	6.6	386	2	H90789	probable aminometh
23	79	6.6	386	2	C85650	probable aminometh
24	79	6.6	437	2	C86823	GTP-binding protei
25	78.5	6.6	188	2	F83816	late competence op
26	78.5	6.6	578	2	F86484	probable hydroxyme
27	78.5	6.6	697	1	QQBEK3	UL104 protein - hu
28	78	6.5	351	2	H82098	UDP-3-O-3-hydroxym
29	77.5	6.5	662	2	T44036	hypothetical prote
30	77.5	6.5	1013	2	T31211	trwC protein homol
31	77.5	6.5	1643	2	T14274	versican precursor
32	77	6.4	381	2	AD1113	hypothetical prote
33	77	6.4	721	2	F87611	TonB-dependent rec
34	77	6.4	738	1	TFHUM	melanotransferrin
35	76.5	6.4	307	2	AG2017	glycerol-3-phospha
36	76.5	6.4	387	2	D69392	probable acyl-CoA
37	76.5	6.4	388	2	S57526	cellulase - Fibrob
38	76.5	6.4	764	2	H98143	cbbBc protein (U60
39	76.5	6.4	764	2	AD3144	formate dehydrogen
40	76.5	6.4	1554	2	T06370	probable DNA (cyto
41	76	6.3	367	2	AH0936	glycerol dehydroge
42	75.5	6.3	310	2	A55053	endothelial monocy
43	75.5	6.3	415	2	T46716	hypothetical prote
44	75.5	6.3	700	2	D70951	probable UvrD - My
45	75.5	6.3	1534	2	S59604	DNA (cytosine-5-)-

ALIGNMENTS

RESULT 1
C86731
copper-potassium transporting ATPase B copB [imported] - Lactococcus lactis subsp. lact
C;Species: Lactococcus lactis subsp. lactis
C;Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 09-Jul-2004
C;Accession: C86731
R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrl
Genome Res. 11, 731-753, 2001
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s
A;Reference number: A86625; MUID:21235186; PMID:11337471
A;Accession: C86731
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-695 <STO>
A;Cross-references: UNIPROT:Q9CH87; UNIPARC:UPI00000C6919; GB:AE005176; PID:g12723778;
A;Experimental source: strain IL1403
C;Genetics:
A;Gene: copB
C;Superfamily: Enterococcus copper-transporting ATPase copB; ATPase nucleotide-binding

Query Match 7.7%; Score 92; DB 2; Length 695;
Best Local Similarity 25.3%; Pred. No. 2;
Matches 38; Conservative 28; Mismatches 50; Indels 34; Gaps 7;

```
Qy      101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
            || ::|            ||  | :| ||     ::   : || | :  || :| : :| |
Db      101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy      161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI-------------QNVPLSE-K 195
            |         |    :: :|:: |||: | :   |:             ::| ||| |
Db      154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy      196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
            : :|   :|  : :: :   ||||   |:|
Db      214 VGDL--LLVKENEKIPADGLILSEALVDES 241
```

RESULT 2
D71293
probable phosphoribosylglycinamide formyltransferase - syphilis spirochete
C;Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C;Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 09-Jul-2004
C;Accession: D71293
R;Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwi
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A;Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A;Reference number: A71250; MUID:98332770; PMID:9665876
A;Accession: D71293
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA

```
A;Residues: 1-597 <COL>
A;Cross-references: UNIPROT:O83693; UNIPARC:UPI00000D3290; GB:AE001243; GB:AE000520; NID
A;Experimental source: strain Nichols
C;Genetics:
A;Gene: TP0695

  Query Match             7.6%;  Score 90.5;  DB 2;  Length 597;
  Best Local Similarity   28.0%;  Pred. No. 2.2;
  Matches   51;  Conservative   19;  Mismatches   47;  Indels   65;  Gaps   12;

Qy         12 CATPPRGDFCGGTERAID----QASFT-------------TSMEWDTQVVKGSSPLGPA 53
              |  |  |      |: |  |     :| ||             |  |||         | | |
Db         90 CALP--GHRLEATKNATDKTRMRACFTRARLRCPRFTFLEPDSFAWDT-------PPGHA 140

Qy         54 GLGAEEPAAGPQLPSWLQP-ERCAVFQC--AQCHAVLADSVHLAWDLSRSLGAVVFSRVT 110
               | :   :||   |  ::| :      |  |||   | ::  :|   |||
Db        141 RLCSHLHSAGLSFPLVVKPTDNMGARGCTLAQCKDTLINACAVARQFSRS---------- 190

Qy        111 NNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMV 170
                |::|  |:|| | ||:|    |:| |:        || |  |||          |: :
Db        191 GRVIIE-EFIVGREFSLEG----LIFDGT--------LYVT--ALA----------DRHI 225

Qy        171 CY 172
               |:
Db        226 CF 227


RESULT 3
T46399
hypothetical protein DKFZp434N2420.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
C;Accession: T46399
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, January 2000
A;Reference number: Z23031
A;Accession: T46399
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-306 <AAA>
A;Cross-references: UNIPROT:Q96EP1; UNIPARC:UPI000006D779; EMBL:AL137561
A;Experimental source: adult testis; clone DKFZp434N2420
C;Genetics:
A;Note: DKFZp434N2420.1

  Query Match             7.4%;  Score 89;  DB 2;  Length 306;
  Best Local Similarity   23.0%;  Pred. No. 1.3;
  Matches   52;  Conservative   26;  Mismatches   74;  Indels   74;  Gaps   11;

Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
              :| : | ||    |       |  :|:  |     | ||:::     ::||  |
Db         76 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 132

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||             ||| |   |    || |  :|
Db        133 PMPDRRAEREQDPRVAPQ-------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 176

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||       | ||   |:  |      : :| |: |:|                | :
Db        177 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 220

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
                :|  || | ||       | ::||    ::    :    ||:| ||
Db        221 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 266


RESULT 4
A43623
kanamycin kinase (EC 2.7.1.95) - Campylobacter jejuni
C;Species: Campylobacter jejuni
C;Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 31-Dec-2004
C;Accession: A43623
R;Tenover, F.C.; Gilbert, T.; O'Hara, P.
Plasmid 22, 52-58, 1989
A;Title: Nucleotide sequence of a novel kanamycin resistance gene, aphA-7, from Campylob
A;Reference number: A43623; MUID:89387451; PMID:2550983
A;Accession: A43623
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-250 <TEN>
A;Cross-references: UNIPROT:P14508; UNIPARC:UPI000012DEBC; GB:M29953; GB:J03316; NID:g14
C;Superfamily: aminoglycoside 3'-phosphotransferase (kanamycin kinase)
C;Keywords: phosphotransferase

  Query Match             7.3%;  Score 88;  DB 2;  Length 250;
  Best Local Similarity   28.7%;  Pred. No. 1.3;
  Matches   33;  Conservative   18;  Mismatches   46;  Indels   18;  Gaps    5;

Qy        105 VFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG-HFC 163
                 :||: | :|  ||  :: :   ||          |  |   :| :  :: ||| |
Db         40 IFSKTTYSVKREAEMMMWLSDKLKVPDV---------IEYGVREHSEYLIMSELRGKHID 90

Qy        164 LSSDKMVCYLLKTKAIVNA----SEMDIQNVPLSEKI-AELKEKIVLTHNRLKSL 213
                 |  : |:   : :|||      :||:| | | ||   |||   |  ||:  :
Db         91 CFIDHPIKYI---ECLVNALHQLQAIDIRNCPFSSKIDVRLKELKYLLDNRIADI 142


RESULT 5
A97229
NAD-dependent DNA ligase [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 31-Dec-2004
C;Accession: A97229
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: A97229
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-669 <KUR>
A;Cross-references: UNIPROT:Q97FQ5; UNIPARC:UPI00000CA5D7; GB:AE001437; PIDN:AAK80620.1,
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2673
C;Superfamily: DNA ligase (NAD), LigA type

  Query Match             7.2%;  Score 86.5;  DB 2;  Length 669;
  Best Local Similarity   27.4%;  Pred. No. 6.2;
  Matches   37;  Conservative   22;  Mismatches   45;  Indels   31;  Gaps    8;

Qy         96 DLSR---SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNL------LFCGSCGIPVGF 146
                |: |    :|: || | :|:|:    | ::|:     :| |  :        :|||  :  |
Db        366 DIGRKKVKIGSRVFVRRSNDVI---PEIMGVTEETEGETNEIEAPTICPYCGSEIVKEGV 422

Qy        147 HLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVP-LSEKIAE-LKEKIV 204
              ||                  ||   : : |      |:||: :  :   |:   ||| || | ||
Db        423 HL-------------FC--ENTLSCKPQMVKSIVHFASREAMNIEGFSEKTAEQLFEK-- 465

Qy        205 LTHNRLKSLMKILSE 219
              |    :  | :|  |
Db        466 LNIKSISDLYRITKE 480


RESULT 6
JQ1977
glutamate synthase (NADH2) (EC 1.4.1.14) precursor [validated] - alfalfa
C;Species: Medicago sativa (alfalfa)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 05-Oct-2004
C;Accession: JQ1977; PQ0551
R;Gregerson, R.G.; Miller, S.S.; Twary, S.N.; Gantt, J.S.; Vance, C.P.
Plant Cell 5, 215-226, 1993
A;Title: Molecular characterization of NADH-dependent glutamate synthase from alfalfa n
A;Reference number: JQ1977; MUID:93200806; PMID:8453303
A;Accession: JQ1977
```

```
A;Molecule type: mRNA
A;Residues: 1-2194 <GRE>
A;Cross-references: UNIPROT:Q03460; UNIPARC:UPI000012B7FA; GB:L01660; NID:g166411; PIDN:
A;Accession: PQ0551
A;Molecule type: protein
A;Residues: 102-114 <GR2>
A;Cross-references: UNIPARC:UPI0000172061
C;Comment: This enzyme catalyzes the reductive transfer of the amido group of glutamine
C;Superfamily: glutamate synthase (NADH/NADPH), eukaryotic type
C;Keywords: 3Fe-4S; chloroplast; iron-sulfur protein; metalloprotein; NAD; oxidoreductas
F;1-101/Domain: propeptide #status predicted <PRO>
F;102-2194/Product: glutamate synthase (NADH) #status experimental <MAT>
F;102/Active site: Cys #status predicted
F;1246,1252,1257/Binding site: 3Fe-4S cluster (Cys) (covalent) #status predicted

  Query Match             7.1%;  Score 85;  DB 1;  Length 2194;
  Best Local Similarity   24.7%;  Pred. No. 39;
  Matches   58;  Conservative   22;  Mismatches   79;  Indels   76;  Gaps   12;

Qy          4 QPLRHRSRCATPPRGDFCGGTERAIDQAS----------FTTSMEWDTQVVKGSSPLGPA 53
              :||   ||   | |          || | :          | : |   :: :|: |
Db       1037 EPLADGSR--NPKRS--------AIKQVASGRFGVSSYYLTNADELQIKMAQGAKP---- 1082

Qy         54 GLGAEEP--------------AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR 99
              | | | |                ||  | |   |    ::        : |    |  ||
Db       1083 GEGGELPGHKVIGDIAITRNSTAGVGLIS--PPPHHDIYS-------IEDLAQLIHDLKN 1133

Qy        100 SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCG--------------SCGIPVG 145
              :  |   :|::  :| ||   |   | :||   ::|  |               | |:|
Db       1134 ANPA---ARISVKLVSEAGVGVIASGVVKGHAEHVLISGHDGGTGASRWTGIKSAGLPWE 1190

Qy        146 FHLYSTHAALAA--LRGHFCLSSDKMVCYLLKT------KAIVNASEMDIQNVPL 192
               |  ||  | |  |||   | :|      |||       |:: | |      ||
Db       1191 LGLAETHQTLVANDLRGRTTLQTDGQ----LKTGRDVAIAALLGAEEYGFSTAPL 1241


RESULT 7
AE2074
ferrichrome binding protein of ABC transporter all2147 [imported] - Nostoc sp. (strain P
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AE2074
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
 Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AE2074
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-361 <KUR>
A;Cross-references: UNIPROT:Q8YV34; UNIPARC:UPI00000CE32E; GB:BA000019; PIDN:BAB73846.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all2147
C;Superfamily: ferrichrome-iron transport protein fecB

  Query Match             7.1%;  Score 84.5;  DB 2;  Length 361;
  Best Local Similarity   26.1%;  Pred. No. 4.3;
  Matches   36;  Conservative   17;  Mismatches   36;  Indels   49;  Gaps    7;

Qy         97 LSRSLGAVVF----SRVT---NNVVLEAPFLVGIEGSLKGSTYNLLFCGSC-----GIPV 144
              :| :|| |       ||     |::|::   :|::          :::|  |     |||
Db         81 ISHALGKVKIPLKPQRVVVLEENIILDSVLALGVK------PVGVMYCQDCEENFRGIP- 133

Qy        145 GFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELKEKIV 204
                                ||     ||    :|     :| | |  |||  ||  ::
Db        134 ---------------------SD-----LLADVPVVG----NIGNQPSLEKILSLKPDLI 163

Qy        205 LTHNRLKSLMKILSEVTP 222
              |    |||  |||| : |
Db        164 LGLTWLKSSYKILSSIAP 181


RESULT 8
D97047
DNA ligase (NAD dependent), LigA [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 31-Dec-2004
C;Accession: D97047
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97047
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-663 <KUR>
A;Cross-references: UNIPROT:Q97JS8; UNIPARC:UPI00000CA10F; GB:AE001437; PIDN:AAK79167.1
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC1195
C;Superfamily: DNA ligase (NAD), LigA type

  Query Match             6.9%;  Score 83;  DB 2;  Length 663;
  Best Local Similarity   28.6%;  Pred. No. 13;
  Matches   36;  Conservative   18;  Mismatches   44;  Indels   28;  Gaps    7;

Qy        102 GAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNL------LFCGSCGIPVGFHLYSTHAAL 155
              || || | :|:|:    | ::|:    :| |  :        :|||  :  | ||
Db        368 GARVFLRRSNDVI---PEIMGVTEETEGETKEIEAPTICPYCGSEIVKEGVHL------- 417

Qy        156 AALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVP-LSEKIAE-LKEKIVLTHNRLKSL 213
                    ||   : : |     |:||: :     |:   ||| || | ||  |    :  |
Db        418 ------FC--ENTLSCKPQMVKSIVHFASRKAMNIEGFSEKTAEQLFEK--LNIKSISDL 467

Qy        214 MKILSE 219
                :|  |
Db        468 YRITKE 473


RESULT 9
B69099
formate hydrogenlyase, iron-sulfur subunit 2 - Methanobacterium thermoautotrophicum (st
C;Species: Methanobacterium thermoautotrophicum
C;Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C;Accession: B69099
R;Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A;Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A;Reference number: A69000; MUID:98037514; PMID:9371463
A;Accession: B69099
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-143 <MTH>
A;Cross-references: UNIPROT:O27769; UNIPARC:UPI0000066781; GB:AE000929; GB:AE000666; NI
A;Experimental source: strain Delta H
C;Genetics:
A;Gene: MTH1736
C;Superfamily: nrfC protein; ferredoxin 2[4Fe-4S] homology
F;62-118/Domain: ferredoxin 2[4Fe-4S] homology <FER2>

  Query Match             6.9%;  Score 82.5;  DB 2;  Length 143;
  Best Local Similarity   23.0%;  Pred. No. 2.1;
  Matches   35;  Conservative   26;  Mismatches   66;  Indels   25;  Gaps    6;

Qy         72 PERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVF------SRVTNNVVLEAPFLVGIEG 125
              || |   :| :|  :   :      |     | ||        |     :     :| ::|
Db          8 PELCD--ECMKCERICPKNAIRVID-----GVPVFCMHCSPERAPCLNICPEDAIVEVDG 60

Qy        126 SLKGSTYNLLFCGSC--GIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNAS 183
              ::       : || |    |||    : :    |::   |:  || :| ::  |  :: |
```

```
Db         61 AVVILEDRCIGCGLCRDACPVG--AITLNERGVAVKCDLCIDRDKPLCVMVCPKGALSES 118

Qy        184 EMDIQNVPLSEKIAELKEKIVLTHNRLKSLMK 215
              |:        :|  ::||     |||:|||
Db        119 SEDM--------MAAKRDKIAGELKRLKNLMK 142


RESULT 10
A85433
sugar transporter like protein [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: A85433
R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin
Nature 402, 769-777, 1999
A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A;Reference number: A85001; MUID:20083488; PMID:10617198
A;Accession: A85433
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-493 <STO>
A;Cross-references: UNIPROT:O23213; UNIPARC:UPI000009F22F; GB:NC_001268; NID:g7270615; P
C;Genetics:
A;Gene: AT4g36670
A;Map position: 4
C;Superfamily: glucose transport protein

  Query Match             6.9%;  Score 82.5;  DB 2;  Length 493;
  Best Local Similarity   24.9%;  Pred. No. 10;
  Matches   53;  Conservative   32;  Mismatches   59;  Indels   69;  Gaps   13;

Qy         79 QCAQCHAVLADSVHLAWDLSRSLGAVVF---SRVTNNVVLEAPFLVGI------EGSLK- 128
              |||   :::  |:   :|     ||:||      ||:| :|   | ||        |||
Db         19 QCAIVASIV--SIIFGYDTGVMSGAMVFIEEDLKTNDVQIEV--LTGILNLCALVGSLLA 74

Qy        129 ----------------------------GSTYNLLFCGSC--GIPVGFHL-----YSTH 152
                                         |  | :|  | |  |: ||| |     ||
Db         75 GRTSDIIGRRYTIVLASILFMLGSILMGWGPNYPVLLSGRCTAGLGVGFALMVAPVYSAE 134

Qy        153 AALAALRG------HFCLSSDKMVCYLLK---TKAIVNAS---EMDIQNVP---LSEKIA 197
               | |: ||       | |:|   :: |::    :|  ::      : |   ||   |:  |
Db        135 IATASHRGLLASLPHLCISIGILLGYIVNYFFSKLPMHIGWRLMLGIAAVPSLVLAFGIL 194

Qy        198 ELKE--KIVLTHNRLKSLMKILSEV--TPDQSK 226
              :: |  : ::   |||   :||  |  :|::::
Db        195 KMPESPRWLIMQGRLKEGKEILELVSNSPEEAE 227


RESULT 11
F87678
DNA polymerase I [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: F87678
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: F87678
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-967 <STO>
A;Cross-references: UNIPROT:Q9A2U2; UNIPARC:UPI00000C7AB1; GB:AE005673; NID:g13425184; P
C;Genetics:
A;Gene: CC3464
C;Superfamily: DNA-directed DNA polymerase I

  Query Match             6.8%;  Score 82;  DB 2;  Length 967;
  Best Local Similarity   24.1%;  Pred. No. 26;
  Matches   51;  Conservative   27;  Mismatches   72;  Indels   62;  Gaps   11;

Qy         12 CATPPRGDFCGGTERAIDQ---ASFTTSMEWDT---QVVKGSSPLGPAGLGAEEPAAGPQ 65
              | ||         | |  |:   |:|   ||: :   :|  ||:   |  |  : ||| |:
Db        271 CDTPLPQPLDALTVREPDKEALAAFLEQMEFRSLARRVGDGSAAATPGTL--DRPAAPPK 328

Qy         66 LP----SWL----------QPERC--AVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSR 108
              |    |::           :| :   | : | :  | |   |  | |     | | |
Db        329 APVVSVSYMGAAARAAAHPVEPVKIDHAAYACVRDLATLKAWVAKATD----KGLVAFDT 384

Qy        109 VTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDK 168
              |:           :|  :|      | ||:           :||   |  |
Db        385 ETD-------------ALSSAT-----AGLCGV----------SLAIAPGEACYIP-- 412

Qy        169 MVCYLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              : :  |   :   :  ||: :||:: || ||
Db        413 -ISHCEKADGLAFEAPADIEQIPLADVIATLK 443


RESULT 12
D90404
transport protein, probable [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004
C;Accession: D90404
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: D90404
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-329 <KUR>
A;Cross-references: UNIPROT:Q97W97; UNIPARC:UPI00000646FA; GB:AE006641; NID:g13815639; P
C;Genetics:
A;Gene: SSO2338

  Query Match             6.8%;  Score 81.5;  DB 2;  Length 329;
  Best Local Similarity   22.2%;  Pred. No. 7.4;
  Matches   40;  Conservative   30;  Mismatches   47;  Indels   63;  Gaps    7;

Qy         87 LADSVHLA-WD---------LSRSLGAVVFSRVTNNVV--------------------LE 116
              |::|:||| |:          | | :|: ::    |:|:                    :
Db         33 LSESMHLAYWEVFAIVALPFLGRIIGSFIYQVFKNSVISYCFPFLGFLVILQNFLGALIF 92

Qy        117 APFLVG----------IEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSS 166
              ||||           :| ::|     |:   : | |:|: :
Db         93 VRFLVGVIFGLLTSYAVESAVKSGRNVLVGFTTAGWPIGWVI----------------- 134

Qy        167 DKMVCY-LLKTKAIVNASEMDIQNVPLSE---KIAELKEKIVLTHNRLKSLMKILSEVTP 222
              | | |||   ::| | : |  : | |    |    : || ::  || |::  :| :||
Db        135 -SYVAYVLLKNWNVINISGILIMLLALFELNGKEFGERSKISVSFPRLTSILIYVSALTP 193


RESULT 13
H85767
glucuronide permease [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C;Accession: H85767
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: H85767
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <STO>
A;Cross-references: UNIPROT:Q8X673; UNIPARC:UPI00000D0BE5; GB:AE005174; NID:g12515601;
A;Experimental source: strain O157:H7, substrain EDL933
C;Genetics:
A;Gene: uidB
```

C;Superfamily: melibiose carrier protein

Query Match 6.8%; Score 81.5; DB 2; Length 457;
Best Local Similarity 25.2%; Pred. No. 11;
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

```
Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              | : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456
```

RESULT 14
C90919
glucuronide permease [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05099
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: C90919
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C90919
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-457 <HAY>
A;Cross-references: UNIPROT:Q8X673; UNIPARC:UPI00000D0BE5; GB:BA000007; PIDN:BAB35746.1;
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2323
C;Superfamily: melibiose carrier protein

Query Match 6.8%; Score 81.5; DB 2; Length 457;
Best Local Similarity 25.2%; Pred. No. 11;
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

```
Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              | : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456
```

RESULT 15
B64918
glucuronide permease uidB - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: B64918
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: B64918
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-457 <BLAT>
A;Cross-references: UNIPROT:P30868; UNIPARC:UPI0000137B23; GB:AE000257; GB:U00096; NID:g
A;Experimental source: strain K-12, substrain MG1655
C;Genetics:
A;Gene: uidB; gusB
C;Superfamily: melibiose carrier protein
C;Keywords: carrier protein; transmembrane protein
F;37-53/Domain: transmembrane #status predicted <TM01>
F;81-97/Domain: transmembrane #status predicted <TM02>
F;152-168/Domain: transmembrane #status predicted <TM03>
F;184-200/Domain: transmembrane #status predicted <TM04>
F;231-247/Domain: transmembrane #status predicted <TM05>
F;263-279/Domain: transmembrane #status predicted <TM06>
F;310-326/Domain: transmembrane #status predicted <TM07>
F;408-424/Domain: transmembrane #status predicted <TM08>

Query Match 6.8%; Score 81.5; DB 2; Length 457;
Best Local Similarity 25.2%; Pred. No. 11;
Matches 37; Conservative 30; Mismatches 53; Indels 27; Gaps 8;

```
Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              | : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456
```

Search completed: May 18, 2006, 15:41:35
Job time : 40 secs

GenCore version 5.1.8

OM protein - protein search, using sw model

Run on: May 18, 2006, 15:33:31 ; Search time 299 Seconds
(without alignments)
708.457 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_7.2:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID		Description
1	1198	100.0	229	1	OIP5_HUMAN	O43482	homo sapien
2	122	10.2	218	2	Q9CXR6_MOUSE	Q9cxr6	mus musculu
3	118	9.8	462	2	Q4PAB5_USTMA	Q4pab5	ustilago ma
4	115.5	9.6	204	2	Q9CZJ6_MOUSE	Q9czj6	m 11 days e
5	113	9.4	233	1	CU045_HUMAN	Q9nyp9	homo sapien
6	113	9.4	233	2	Q542Z0_HUMAN	Q542z0	homo sapien
7	110.5	9.2	232	1	CU045_PANTR	Q68ut5	pan troglod
8	109	9.1	207	2	Q5BLB1_BRARE	Q5blb1	brachydanio
9	107	8.9	532	2	Q54HI6_DICDI	Q54hi6	dictyosteli
10	105	8.8	155	1	YCZC_SCHPO	Q9p802	schizosacch
11	96	8.0	391	2	Q33BJ5_ORYSA	Q33bj5	oryza sativ
12	96	8.0	497	2	Q3V0J4_MOUSE	Q3v0j4	mus musculu
13	94	7.8	1556	2	Q4Q626_LEIMA	Q4q626	leishmania
14	93.5	7.8	702	2	Q2QSG3_ORYSA	Q2qsg3	oryza sativ
15	92	7.7	361	2	Q8S611_ORYSA	Q8s611	oryza sativ
16	92	7.7	695	2	Q9CH87_LACLA	Q9ch87	lactococcus
17	91.5	7.6	410	2	Q2JZY7_RHIET	Q2jzy7	rhizobium e
18	91	7.6	628	2	Q69KE0_ORYSA	Q69ke0	oryza sativ
19	90.5	7.6	272	2	Q40QJ4_DESAC	Q40qj4	desulfuromo
20	90.5	7.6	597	2	O83693_TREPA	O83693	treponema p
21	89.5	7.5	502	2	Q50PD2_ENTHI	Q50pd2	entamoeba h
22	89.5	7.5	661	2	Q50V43_ENTHI	Q50v43	entamoeba h
23	89.5	7.5	661	2	Q50WZ8_ENTHI	Q50wz8	entamoeba h
24	89.5	7.5	2034	2	Q4E556_TRYCR	Q4e556	trypanosoma
25	89	7.4	633	2	Q5M052_STRT1	Q5m052	streptococc
26	89	7.4	664	1	CHFR_HUMAN	Q96ep1	homo sapien
27	89	7.4	735	2	Q5M4R3_STRT2	Q5m4r3	streptococc
28	89	7.4	893	1	POL2_BAMMA	Q65329	barley mild
29	88.5	7.4	152	2	Q50XW7_ENTHI	Q50xw7	entamoeba h
30	88.5	7.4	384	2	Q3A3W4_PELCD	Q3a3w4	pelobacter
31	88.5	7.4	638	2	Q75RZ6_CIOIN	Q75rz6	ciona intes
32	88	7.3	250	1	KKA7_CAMJE	P14508	campylobact
33	88	7.3	493	2	Q68VH2_9POTY	Q68vh2	barley mild
34	87.5	7.3	634	2	Q6CX82_KLULA	Q6cx82	kluyveromyc
35	87.5	7.3	663	2	Q3U233_MOUSE	Q3u233	m nod-deriv
36	87.5	7.3	696	2	Q4JTS4_CORJK	Q4jts4	corynebacte
37	87	7.3	524	2	Q3TFD8_MOUSE	Q3tfd8	mus musculu
38	87	7.3	592	2	Q3UGJ9_MOUSE	Q3ugj9	mus musculu
39	87	7.3	664	1	CHFR_MOUSE	Q8101l3	mus musculu
40	87	7.3	664	2	Q3U4U9_MOUSE	Q3u4u9	mus musculu
41	86.5	7.2	669	2	Q97FQ5_CLOAB	Q97fq5	clostridium
42	86.5	7.2	2093	2	Q5K2K3_9RHAB	Q5k2k3	isfahan vir
43	86	7.2	145	2	Q839T4_ENTFA	Q839t4	enterococcu
44	86	7.2	327	2	Q4NE64_9MICC	Q4ne64	arthrobacte
45	86	7.2	331	2	Q6GXE5_COCCA	Q6gxe5	cochliobolu

ALIGNMENTS

```
RESULT 1
OIP5_HUMAN
ID   OIP5_HUMAN     STANDARD;      PRT;   229 AA.
AC   O43482; Q96BX7;
DT   03-APR-2002, integrated into UniProtKB/Swiss-Prot.
DT   03-APR-2002, sequence version 2.
DT   07-FEB-2006, entry version 29.
DE   Opa-interacting protein 5.
GN   Name=OIP5;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [MRNA].
RX   MEDLINE=98125741; PubMed=9466265;
RA   Williams J.M., Chen G.-C., Zhu L., Rest R.F.;
RT   "Using the yeast two-hybrid system to identify human epithelial cell
RT   proteins that bind gonococcal Opa proteins: intracellular gonococci
RT   bind pyruvate kinase via their Opa proteins and require host pyruvate
RT   for growth.";
RL   Mol. Microbiol. 27:171-186(1998).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC   TISSUE=Uterus;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SUBUNIT: Binds outer membrane protein OpaP from Neisseria
CC       gonorrhoeae.
CC   -!- INTERACTION:
CC       P04049:RAF1; NbExp=3; IntAct=EBI-536879, EBI-365996;
CC   ----------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
```

```
CC   ----------------------------------------------------------------------
DR   EMBL; AF025441; AAC39561.1; ALT_INIT; mRNA.
DR   EMBL; BC015050; AAH15050.1; -; mRNA.
DR   IntAct; O43482; -.
DR   Ensembl; ENSG00000104147; Homo sapiens.
DR   H-InvDB; HIX0012150; -.
DR   HGNC; HGNC:20300; OIP5.
DR   MIM; 606020; gene.
DR   LinkHub; O43482; -.
DR   GO; GO:0005515; F:protein binding; TAS.
DR   GO; GO:0007154; P:cell communication; NAS.
FT   CHAIN         1    229       Opa-interacting protein 5.
FT                                /FTId=PRO_0000058038.
SQ   SEQUENCE   229 AA;  24691 MW;  0EBD4006193A3106 CRC64;

  Query Match             100.0%;  Score 1198;  DB 1;  Length 229;
  Best Local Similarity   100.0%;  Pred. No. 1.3e-104;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229


RESULT 2
Q9CXR6_MOUSE
ID   Q9CXR6_MOUSE   PRELIMINARY;   PRT;   218 AA.
AC   Q9CXR6;
DT   01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT   01-JUN-2001, sequence version 1.
DT   07-FEB-2006, entry version 18.
DE   13 days embryo head cDNA, RIKEN full-length enriched library,
DE   clone:3110025H23 product:hypothetical protein, full insert sequence.
GN   Name=2610039C10Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Head; STRAIN=C57BL/6J;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Methods Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   PubMed=16141072; DOI=10.1126/science.1112014;
RA   Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA   Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA   Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA   Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA   Ambesi-Impiombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA   Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA   Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA   Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA   di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
RA   Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA   Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA   Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA   Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA   Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA   Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA   Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA   Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA   Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA   Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
RA   Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA   Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
RA   Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA   Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA   Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA   Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA   Sperling S., Stupka E., Sugiura K., Sultana R., Takenaka Y., Taki K.,
RA   Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA   Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA   Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA   Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA   Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA   Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA   Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA   Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA   Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA   Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA   Hayashizaki Y.;
RT   "The transcriptional landscape of the mammalian genome.";
RL   Science 309:1559-1563(2005).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   PubMed=16141073; DOI=10.1126/science.1112009;
RG   RIKEN Genome Exploration Research Group, and Genome Science Group
RG   (Genome Network Core Team) and the FANTOM Consortium;
RT   "Antisense Transcription in the Mammalian Transcriptome.";
RL   Science 309:1564-1566(2005).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
```

```
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [7]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [8]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Head;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AK014084; BAB29147.1; -; mRNA.
DR   Ensembl; ENSMUSG00000022978; Mus musculus.
DR   MGI; MGI:1913828; 2610039C10Rik.
KW   Hypothetical protein.
SQ   SEQUENCE   218 AA;  24447 MW;  55390E5BFDDDE8C2 CRC64;

  Query Match             10.2%;  Score 122;  DB 2;  Length 218;
  Best Local Similarity   23.3%;  Pred. No. 0.0065;
  Matches   48;  Conservative   30;  Mismatches  102;  Indels   26;  Gaps    6;

Qy         29 DQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLA 88
              |  :  :    :|    :  :  ||      || ||  : |        || ||:|   |
Db         15 DSSRYLRLQKWAN--MSSADALGLEKERPEEKAAAAENP--------LVFLCARCRRPLG 64
..
Qy         89 DSVHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFH 147
              ||  | |  |:      ::   |: || ::    :        |     |:|  | : :|:
Db         65 DS--LTWVASQEDTNCILLRSVSCNVSVDKEPKLSKCRDEDGCILEALYCTGCSLSLGYV 122

Qy        148 LYSTHAALAALRGHFCLSSDKMVCYLL---------KTKAIVNASEMDIQNVPLSEKIAE 198
                |   |   |  |||| : :  | |          | | : |        | : : | :
Db        123 YRCTPKNLDYKRDLFCLSVETVESYTLGSSEKQIVSKDKELFNLE----SRVEIEKSIKQ 178

Qy        199 LKEKIVLTHNRLKSLMKILSEVTPDQ 224
              ::| :     :|: :   ||   | |
Db        179 MEEVLTALQKKLREVESKLSLAQPGQ 204


RESULT 3
Q4PAB5_USTMA
ID   Q4PAB5_USTMA   PRELIMINARY;   PRT;   462 AA.
AC   Q4PAB5;
DT   19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT   19-JUL-2005, sequence version 1.
DT   07-FEB-2006, entry version 3.
DE   Hypothetical protein.
GN   ORFNames=UM02948.1;
OS   Ustilago maydis 521.
OC   Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC   Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX   NCBI_TaxID=237631;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=521;
RA   Birren B.W., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA   Ait-Zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA   Arachchi H.M., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA   Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA   Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA   Calvo S.E., Camarata J., Campo K., Chang J., Cheshatsang Y.,
RA   Citroen M., Collymore A., Considine T., Cook A., Cooke P., Corum B.,
RA   Cuomo C., David R., Dawoe T., Degray S., Dodge S., Dooley K.,
RA   Dorje P., Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T.,
RA   Engels R., Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA   Fitzgerald M., Foley K., Gage D., Galagan J.E., Gearin G., Gnerre S.,
RA   Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA   Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA   Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA   Jaffe D., Jones C., Kamal M., Kamat A., Kamvysselis M., Karlsson E.,
RA   Kells C., Kieu A., Kisner P., Kodira C., Kulbokas E., Labutti K.,
RA   Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA   Lindblad-Toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA   Lui A., Ma L.-J., Mabbitt R., MacDonald J., MacLean C., Major J.,
RA   Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA   McCarthy M., McDonough S., McGhee T., Meldrim J., Meneus L.,
RA   Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA   Mozes J., Mulrain L., Munson G., Naylor J., Newes C., Nguyen C.,
RA   Nguyen N., Nguyen T., Nicol R., Nielsen C.B., Nizzari M., Norbu C.,
RA   Norbu N., O'Donnell P., Okoawo O., O'Leary S., Omotosho B.,
RA   O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piqani B.,
RA   Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA   Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA   Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA   Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA   Spencer B., Stalker J., Stange-Thomann N., Stavropoulos S.,
RA   Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA   Tenzing P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA   Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA   Venkataraman V.S., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA   Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA   Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA   Zimmer A., Zody M., Lander E.S.;
RT   "The genome sequence of Ustilago maydis.";
RL   Submitted (FEB-2004) to the EMBL/GenBank/DDBJ databases.
CC   -!- CAUTION: The sequence shown here is derived from an
```

```
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AACP01000101; EAK84120.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   462 AA;  49283 MW;  EDDCFDF17757B509 CRC64;

  Query Match             9.8%;  Score 118;  DB 2;  Length 462;
  Best Local Similarity   25.9%;  Pred. No. 0.04;
  Matches   64;  Conservative   28;  Mismatches  109;  Indels   46;  Gaps   10;

Qy          2 AAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQV-VKGSSPLGPAGLGAEEP 60
              |||||        ||:|              | ||:    |:   :|     | | |
Db        109 AAQPLSSDDDDMAPPKG------------GSLTTT---STRARGRGRGRGRPRGRGRG 153

Qy         61 AA-------GPQL--PSWLQPERCAVFQCAQCHAVLADSV-HLAWDLSRSLGAVVFSRVT 110
              |:        | |       :|    ||||  |  :| ||:  :| |:   || |: | |:
Db        154 ASTSTTVVHKPDLEESDSEEPPNPLVFQCRCCFRLLGDSLAFVATDI--DLGYVILSDVS 211

Qy        111 NNVVLEAPFLVGIE-GSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKM 169
               :  :  :    | |   |||:  | | |   || :  :|   |  ||  | |  | :
Db        212 EIIQQDTTYETSTEPGKDIGSTFARLRCAGCNAAVGRNYRTTPRDLDDLRDCFSLEVDAI 271

Qy        170 VCYLLKTKAIVNASEMDIQ---------------NVPLSEKIAELKEKIVLTH-NRLKS 212
              | | :       | |                  : |  | :     |:::|   | ::
Db        272 YTYQLGSNYTRQKEEDDEDQDALVDGGASGKPARASRPTHETVQAEDTKVLITKMERTRA 331

Qy        213 LMKILSE 219
              |   ||:
Db        332 LTIELSD 338


RESULT 4
Q9CZJ6_MOUSE
ID   Q9CZJ6_MOUSE   PRELIMINARY;   PRT;   204 AA.
AC   Q9CZJ6;
DT   01-JUN-2001, integrated into UniProtKB/TrEMBL.
DT   01-JUN-2001, sequence version 1.
DT   07-FEB-2006, entry version 21.
DE   11 days embryo whole body cDNA, RIKEN full-length enriched library,
DE   clone:2700078I24 product:RIKEN cDNA 2610039C10 (2 cells egg cDNA,
DE   RIKEN full-length enriched library, clone:B020013B09
DE   product:hypothetical protein, full insert sequence) (2610039C10Rik
DE   protein).
GN   Name=2610039C10Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg, and Whole body;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Methods Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg, and Whole body;
RX   PubMed=16141072; DOI=10.1126/science.1112014;
RA   Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA   Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA   Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA   Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA   Ambesi-Impiombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA   Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA   Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA   Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA   di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
RA   Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA   Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA   Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA   Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA   Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA   Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA   Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA   Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA   Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA   Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
RA   Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA   Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
RA   Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA   Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA   Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA   Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA   Sperling S., Stupka E., Sugiura K., Sultana R., Takenaka Y., Taki K.,
RA   Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA   Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA   Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA   Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA   Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA   Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA   Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA   Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA   Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA   Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA   Hayashizaki Y.;
RT   "The transcriptional landscape of the mammalian genome.";
RL   Science 309:1559-1563(2005).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg, and Whole body;
RX   PubMed=16141073; DOI=10.1126/science.1112009;
RG   RIKEN Genome Exploration Research Group, and Genome Science Group
RG   (Genome Network Core Team) and the FANTOM Consortium;
RT   "Antisense Transcription in the Mammalian Transcriptome.";
RL   Science 309:1564-1566(2005).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg, and Whole body;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
```

```
RC   STRAIN=C57BL/6J; TISSUE=Egg, and Whole body;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg, and Whole body;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
RL   Genome Res. 10:1617-1630(2000).
RN   [7]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg, and Whole body;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [8]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Whole body;
RA   Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
RA   Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA   Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA   Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA   Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
RA   Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA   Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA   Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA   Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA   Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
RN   [9]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Egg;
RA   Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA   Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,
RA   Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
RA   Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA   Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (APR-2004) to the EMBL/GenBank/DDBJ databases.
RN   [10]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6; TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

  Query Match             9.6%;  Score 115.5;  DB 2;  Length 204;
  Best Local Similarity   25.6%;  Pred. No. 0.024;
  Matches   50;  Conservative   29;  Mismatches   97;  Indels   19;  Gaps    7;

Qy         29 DQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLA 88
              | : :    :|    :  :  ||      || ||  : |         || ||:|   |
Db         15 DSSRYLRLQKWAN--MSSADALGLEKERPEEKAAAAENP--------LVFLCARCRRPLG 64

Qy         89 DSVHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFH 147
              ||  | |  |:     ::   |: || ::   :       |     |:|  | : :|:
Db         65 DS--LTWVASQEDTNCILLRSVSCNVSVDKEPKLSKCRDEDGCILEALYCTGCSLSLGYV 122

Qy        148 LYSTHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVL 205
                |   |   | |||| : :  | |   | || : : :: |:   |   |::: |
Db        123 YRCTPKNLDYKRDLFCLSVEAVESYTLGSSEKQIV-SEDKELFNL---ESRVEIEKSIKQ 178

Qy        206 THNRLKSLMKILSEV 220
                | :| | | ||
Db        179 MEEVLTALQKKLREV 193


RESULT 5
CU045_HUMAN
ID   CU045_HUMAN    STANDARD;      PRT;   233 AA.
AC   Q9NYP9;
DT   11-JAN-2001, integrated into UniProtKB/Swiss-Prot.
DT   01-OCT-2000, sequence version 1.
DT   07-FEB-2006, entry version 27.
DE   Protein C21orf45.
GN   Name=C21orf45;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [MRNA].
RX   MEDLINE=20237674; PubMed=10773462; DOI=10.1016/S0378-1119(00)00089-5;
RA   Slavov D., Hattori M., Sakaki Y., Rosenthal A., Shimizu N.,
RA   Minoshima S., Kudoh J., Yaspo M.-L., Ramser J., Reinhardt R.,
RA   Reimer C., Clancy K., Rynditch A., Gardiner K.;
RT   "Criteria for gene identification and features of genome organization:
RT   analysis of 6.5 Mb of DNA sequence from human chromosome 21.";
RL   Gene 247:215-232(2000).
RN   [2]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE MRNA].
RC   TISSUE=Blood;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
```

```
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- SIMILARITY: To S.pombe SpC970.12.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AF231921; AAF72945.1; -; mRNA.
DR   EMBL; BC042917; AAH42917.1; -; mRNA.
DR   Ensembl; ENSG00000159055; Homo sapiens.
DR   HGNC; HGNC:1286; C21orf45.
FT   CHAIN         1    233       Protein C21orf45.
FT                                /FTId=PRO_0000079515.
SQ   SEQUENCE   233 AA;  25863 MW;  A07522806C4B6221 CRC64;

  Query Match             9.4%;  Score 113;  DB 1;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.05;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 6
Q542Z0_HUMAN
ID   Q542Z0_HUMAN   PRELIMINARY;   PRT;   233 AA.
AC   Q542Z0;
DT   24-MAY-2005, integrated into UniProtKB/TrEMBL.
DT   24-MAY-2005, sequence version 1.
DT   07-MAR-2006, entry version 8.
DE   Hypothetical protein FLJ90800 (FAPP1-associated protein 1).
GN   Name=FASP1;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RX   PubMed=16303743; DOI=10.1093/dnares/12.2.117;
RA   Otsuki T., Ota T., Nishikawa T., Hayashi K., Suzuki Y., Yamamoto J.,
RA   Wakamatsu A., Kimura K., Sakamoto K., Hatano N., Kawai Y., Ishii S.,
RA   Saito K., Kojima S., Sugiyama T., Ono T., Okano K., Yoshikawa Y.,
RA   Aotsuka S., Sasaki N., Hattori A., Okumura K., Nagai K., Sugano S.,
RA   Isogai T.;
RT   "Signal sequence and keyword trap in silico for selection of full-
RT   length human cDNAs encoding secretion or membrane proteins from oligo-
RT   capped cDNA libraries.";
RL   DNA Res. 12:117-126(2005).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=21061410; PubMed=11001876; DOI=10.1042/0264-6021:3510019;
RA   Dowler S.J., Currie R.A., Campbell D.G., Deak M., Kular G.,
RA   Downes C.P., Alessi D.R.;
RT   "Identification of pleckstrin-homology-domain-containing proteins with
RT   novel phosphoinositide-binding specificities.";
RL   Biochem. J. 351:19-31(2000).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   TISSUE=Brain;
RA   Dowler S.J.;
RT   "Identification of FAPP1 interacting proteins.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AK075281; BAC11517.1; -; mRNA.
DR   EMBL; AF387845; AAK70498.1; -; mRNA.
DR   Ensembl; ENSG00000159055; Homo sapiens.
SQ   SEQUENCE   233 AA;  25863 MW;  A07522806C4B6221 CRC64;

  Query Match             9.4%;  Score 113;  DB 2;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.05;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221


RESULT 7
CU045_PANTR
ID   CU045_PANTR    STANDARD;      PRT;   232 AA.
AC   Q68UT5;
DT   23-NOV-2004, integrated into UniProtKB/Swiss-Prot.
DT   11-OCT-2004, sequence version 1.
DT   07-FEB-2006, entry version 9.
DE   Protein C21orf45 homolog.
OS   Pan troglodytes (Chimpanzee).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae;
OC   Pan.
OX   NCBI_TaxID=9598;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX   PubMed=15164055; DOI=10.1038/nature02564;
RA   Watanabe H., Fujiyama A., Hattori M., Taylor T.D., Toyoda A.,
RA   Kuroki Y., Noguchi H., BenKahla A., Lehrach H., Sudbrak R., Kube M.,
RA   Taenzer S., Galgoczy P., Platzer M., Scharfe M., Nordsiek G.,
RA   Bloecker H., Hellmann I., Khaitovich P., Paabo S., Reinhardt R.,
RA   Zheng H.-J., Zhang X.-L., Zhu G.-F., Wang B.-F., Fu G., Ren S.-X.,
RA   Zhao G.-P., Chen Z., Lee Y.-S., Cheong J.-E., Choi S.-H., Wu K.-M.,
RA   Liu T.-T., Hsiao K.-J., Tsai S.-F., Kim C.-G., Oota S., Kitano T.,
RA   Kohara Y., Saitou N., Park H.-S., Wang S.-Y., Yaspo M.-L., Sakaki Y.;
RT   "DNA sequence and comparative analysis of chimpanzee chromosome 22.";
RL   Nature 429:382-388(2004).
CC   -!- SIMILARITY: To S.pombe SpC970.12.
```

```
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AL954205; CAH18576.1; -; Genomic_DNA.
FT   CHAIN       1    232       Protein C21orf45 homolog.
FT                              /FTId=PRO_0000079516.
SQ   SEQUENCE   232 AA;  25832 MW;  6549B8BE369245E7 CRC64;

  Query Match           9.2%;  Score 110.5;  DB 1;  Length 232;
  Best Local Similarity   27.1%;  Pred. No. 0.086;
  Matches   52;  Conservative   29;  Mismatches   88;  Indels   23;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||          ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERARL--EEAAAA--------EERPLVFLCSGCRRPLGDS 94

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :        |     | |  | : :|:
Db         95 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 152

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   | |||| : :  |:|    | || : : :: |:   |   |::: :
Db        153 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 208

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        209 DVLKALQMKLWE 220
```

```
RESULT 8
Q5BLB1_BRARE
ID   Q5BLB1_BRARE   PRELIMINARY;   PRT;   207 AA.
AC   Q5BLB1;
DT   12-APR-2005, integrated into UniProtKB/TrEMBL.
DT   12-APR-2005, sequence version 1.
DT   07-FEB-2006, entry version 6.
DE   LOC553502 protein (Fragment).
GN   Name=LOC553502;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Singapore local strain; TISSUE=Embryo;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Singapore local strain; TISSUE=Embryo;
RG   NIH MGC Project;
RL   Submitted (FEB-2005) to the EMBL/GenBank/DDBJ databases.
```

```
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; BC090521; AAH90521.1; -; mRNA.
DR   Ensembl; ENSDARG00000035333; Danio rerio.
FT   NON_TER       1      1
SQ   SEQUENCE   207 AA;  22941 MW;  AF1A0ACCF375A28C CRC64;

  Query Match           9.1%;  Score 109;  DB 2;  Length 207;
  Best Local Similarity   25.5%;  Pred. No. 0.1;
  Matches   37;  Conservative   28;  Mismatches   66;  Indels   14;  Gaps    6;

Qy         76 AVFQCAQCHAVLADSVHLAWDLS-RSLGAVVFSRVTNNVVL-EAPFLVGIEGSLKGSTYN 133
              ||| | :|   : ||  |:|  |       ::  |:|:|:|: : ||: |     |
Db         54 AVFMCGKCKLPIGDS--LSWAGSDDENNQIMLKRITDNIVVGKEPFVSGTRKELGCLVVN 111

Qy        134 LLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLS 193
               |  |  |   :|    ||   |   |  || : : :  |:: :       |:| :: |::
Db        112 LT-CRGCCSELGVMYISTPKKLDCKRSLFCFNVENIESYVVGSPG-QQMPELDREDKPVT 169

Qy        194 --------EKIAELKEKIVLTHNRL 210
                     ::: |:    |:   ||
Db        170 LEYQDTVHQQMTEITSLAVIIGQRL 194
```

```
RESULT 9
Q54HI6_DICDI
ID   Q54HI6_DICDI   PRELIMINARY;   PRT;   532 AA.
AC   Q54HI6;
DT   24-MAY-2005, integrated into UniProtKB/TrEMBL.
DT   24-MAY-2005, sequence version 1.
DT   07-FEB-2006, entry version 3.
DE   Hypothetical protein.
GN   ORFNames=DDB0188417;
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=AX4;
RX   PubMed=15875012; DOI=10.1038/nature03481;
RA   Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.A.,
RA   Sucgang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,
RA   Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,
RA   Bankier A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,
RA   Pilcher K., Chen G., Saunders D., Sodergren E.J., Davis P.,
RA   Kerhornou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,
RA   Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.M.,
RA   Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,
RA   Muzny D.M., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,
RA   Hauser H., James K.D., Quiles M., Madan Babu M., Saito T.,
RA   Buchrieser C., Wardroper A., Felder M., Thangavelu M., Johnson D.,
RA   Knights A., Loulseged H., Mungall K.L., Oliver K., Price C.,
RA   Quail M.A., Urushihara H., Hernandez J., Rabbinowitsch E., Steffen D.,
RA   Sanders M., Ma J., Kohara Y., Sharp S., Simmonds M.N., Spiegler S.,
RA   Tivey A., Sugano S., White B., Walker D., Woodward J.R., Winckler T.,
RA   Tanaka Y., Shaulsky G., Schleicher M., Weinstock G.M., Rosenthal A.,
RA   Cox E.C., Chisholm R.L., Gibbs R.A., Loomis W.F., Platzer M.,
RA   Kay R.R., Williams J.G., Dear P.H., Noegel A.A., Barrell B.G.,
RA   Kuspa A.;
RT   "The genome of the social amoeba Dictyostelium discoideum.";
RL   Nature 435:43-57(2005).
CC   -!- CAUTION: The sequence shown here is derived from an
CC       EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC       preliminary data.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AAFI01000207; EAL62738.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   532 AA;  60715 MW;  BAF75AB8CD7C1D5C CRC64;
```

```
  Query Match             8.9%;  Score 107;  DB 2;  Length 532;
  Best Local Similarity   22.9%;  Pred. No. 0.53;
  Matches   38;  Conservative   29;  Mismatches   75;  Indels   24;  Gaps    5;

Qy         77 VFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFLVGIEGSLKGST----- 131
              :| |  |   : ||  :   :::|    ::  :||  |::       | |  | ||
Db        205 IFSCLSCKTPIGDSTLI---VNKSHECDMYMTLTNTFVMKKLGGSVILGDQKPSTYRGET 261

Qy        132 ---YNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLL-KTKAIVNASEMDI 187
                 |:| :|  |  |:|      |      | || :   :  | :     |:   : :|
Db        262 DCIYSLFYCSYCNSPLGRKYNKTSQKFKILNNHFVVDIGSITYYSVGGNNNIIVEDKNEI 321

Qy        188 QNVPLS----------EKIAELKEKIVLTHNRLKSLMKILSEVTPD 223
                ||||          | : :: |||    ::| |:   : :  |:
Db        322 --VPLSFGNDKLTKMHETMMKMNEKINFLEDQLLSVFNSIKKALPE 365


RESULT 10
YCZC_SCHPO
ID   YCZC_SCHPO     STANDARD;      PRT;   155 AA.
AC   Q9P802;
DT   11-JAN-2001, integrated into UniProtKB/Swiss-Prot.
DT   01-OCT-2000, sequence version 1.
DT   07-FEB-2006, entry version 21.
DE   Hypothetical protein C970.12 in chromosome III.
GN   ORFNames=SPCC970.12;
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360; DOI=10.1038/nature724;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M.H., Lyne R., Stewart A.,
RA   Sgouros J.G., Peat N., Hayles J., Baker S.G., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D.E., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K.D., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K.L., Murphy L.D., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K.M., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M.N., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J.R., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Duesterhoeft A., Fritzc C., Holzer E., Moestl D.,
RA   Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R.,
RA   Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- SIMILARITY: To human C21orf45.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AL031530; CAB72327.1; -; Genomic_DNA.
DR   GeneDB_Spombe; SPCC970.12; -.
DR   BioCyc; SPOM-XXX-01:SPOM-XXX-01-002203-MONOMER; -.
DR   GO; GO:0031511; C:Mis6 centromere subcomplex; TAS.
DR   GO; GO:0005515; F:protein binding; IPI.
DR   GO; GO:0031055; P:chromatin remodeling at centromere; IMP.
DR   GO; GO:0007059; P:chromosome segregation; IMP.
DR   GO; GO:0016575; P:histone deacetylation; IMP.
DR   GO; GO:0051382; P:kinetochore assembly; TAS.
DR   GO; GO:0031503; P:protein complex localization; TAS.
DR   GO; GO:0031066; P:regulation of histone deacetylation at cent. . .; IMP.
KW   Complete proteome; Hypothetical protein.
FT   CHAIN         1    155       Hypothetical protein C970.12 in
FT                                chromosome III.
FT                                /FTId=PRO_0000116562.
SQ   SEQUENCE   155 AA;  17874 MW;  A7AEBD0F93760B0B CRC64;

  Query Match             8.8%;  Score 105;  DB 1;  Length 155;
  Best Local Similarity   30.1%;  Pred. No. 0.17;
  Matches   41;  Conservative   18;  Mismatches   61;  Indels   16;  Gaps    6;

Qy         71 QPERCAVFQCAQCHAVLADSVHLAWDLSRS--LGAVVFSRVTNNVVLEAPFLVGIEGSLK 128
              ||   :|||| :|  :: ||   || :|    |   :   | |:| :|  |    :|
Db         19 QP---SVFQCKKCFQIVGDS--NAWVISHREYLSFTLSDAVENSVRVEDTFKRSDDGL-- 71

Qy        129 GSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQ 188
              |: | |  |   :|    ||   |  :|  :  | ||:  | |  |  ||   :
Db         72 -CVYSELSCTRCNEVIGKVYNSTPIYLDDIRDMYTFSMDKLQAYQLGNKT-VNPEGLTRY 129

Qy        189 NVPLSEKIAELKEKIV 204
               | |     |::| |:
Db        130 QVDL-----EMREDII 140


RESULT 11
Q33BJ5_ORYSA
ID   Q33BJ5_ORYSA   PRELIMINARY;   PRT;   391 AA.
AC   Q33BJ5;
DT   06-DEC-2005, integrated into UniProtKB/TrEMBL.
DT   06-DEC-2005, sequence version 1.
DT   07-FEB-2006, entry version 3.
DE   Hypothetical protein.
GN   ORFNames=LOC_Os10g01330;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Buell C.R., Wing R.A., McCombie W.R., Messing J., Yuan Q., Ouyang S.;
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AE016959; ABB46574.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   391 AA;  42165 MW;  F2DE86B4D4C41FBD CRC64;

  Query Match             8.0%;  Score 96;  DB 2;  Length 391;
  Best Local Similarity   26.1%;  Pred. No. 3.9;
  Matches   69;  Conservative   29;  Mismatches   90;  Indels   76;  Gaps   16;

Qy          2 AAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPA 61
              |  ||  |          |  |||  |    : |     |  || ||    |:|  | :|
Db         49 APTPLPFRRAVRAKAAQDGSGGTSSASPAVAST-----DVVVVPGSREATPSG-PASDPV 102

Qy         62 AGPQLP----SW--LQPERCAVFQC----------AQCHAVLAD--SVHLAWDLS---- 98
              ||   |    ||  || |   : :            |:  || |:  :  || ||:
Db        103 AGRGSPAAVLSWEELQVEMGRLLEAGARVIGREIAEARVAAVSANKRADRLAHDLAEARE 162

Qy         99 -----RSLGAV-VFSRVTNNVVLEAPFLVGIEGSLKGSTYNLL--FCGSCGI-------P 143
                  | ||     | : ||       |  | |||:  ||  |    | |||       |
Db        163 DLKKMRELGLEHRMSELGNN-------LSEIRGSLR-VTYTGLHQLAGKCGIKSTIPANP 214

Qy        144 VGFHLYSTHAALAAL-----RGHFCLSSDKM-------VCYLLKTKAIVNASEMDIQNV- 190
                | | |: | |||         |      ::|        |::|    : :  |:|:: :
Db        215 DEFSLTSSLAELAAAMEEIPSKHAARIGEEMSNRIYIGACHILACVRLAH-PELDLREIL 273
```

```
Qy        191 ----------PLSEKIAELKEKIV 204
                         : |:: :| : ::
Db        274 DQGEASDARKDVMEEVGDLGKSVL 297


RESULT 12
Q3V0J4_MOUSE
ID   Q3V0J4_MOUSE   PRELIMINARY;   PRT;   497 AA.
AC   Q3V0J4;
DT   11-OCT-2005, integrated into UniProtKB/TrEMBL.
DT   11-OCT-2005, sequence version 1.
DT   07-FEB-2006, entry version 6.
DE   Adult male testis cDNA, RIKEN full-length enriched library,
DE   clone:4930572N15 product:hypothetical Ankyrin/Ankyrin repeat
DE   profile/Ankyrin repeat region circular profile containing protein,
DE   full insert sequence.
GN   Name=4930564N15Rik;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC   Muroidea; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=99279253; PubMed=10349636; DOI=10.1016/S0076-6879(99)03004-9;
RA   Carninci P., Hayashizaki Y.;
RT   "High-efficiency full-length cDNA cloning.";
RL   Methods Enzymol. 303:19-44(1999).
RN   [2]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   PubMed=16141072; DOI=10.1126/science.1112014;
RA   Carninci P., Kasukawa T., Katayama S., Gough J., Frith M.C., Maeda N.,
RA   Oyama R., Ravasi T., Lenhard B., Wells C., Kodzius R., Shimokawa K.,
RA   Bajic V.B., Brenner S.E., Batalov S., Forrest A.R., Zavolan M.,
RA   Davis M.J., Wilming L.G., Aidinis V., Allen J.E.,
RA   Ambesi-Impiombato A., Apweiler R., Aturaliya R.N., Bailey T.L.,
RA   Bansal M., Baxter L., Beisel K.W., Bersano T., Bono H., Chalk A.M.,
RA   Chiu K.P., Choudhary V., Christoffels A., Clutterbuck D.R.,
RA   Crowe M.L., Dalla E., Dalrymple B.P., de Bono B., Della Gatta G.,
RA   di Bernardo D., Down T., Engstrom P., Fagiolini M., Faulkner G.,
RA   Fletcher C.F., Fukushima T., Furuno M., Futaki S., Gariboldi M.,
RA   Georgii-Hemming P., Gingeras T.R., Gojobori T., Green R.E.,
RA   Gustincich S., Harbers M., Hayashi Y., Hensch T.K., Hirokawa N.,
RA   Hill D., Huminiecki L., Iacono M., Ikeo K., Iwama A., Ishikawa T.,
RA   Jakt M., Kanapin A., Katoh M., Kawasawa Y., Kelso J., Kitamura H.,
RA   Kitano H., Kollias G., Krishnan S.P., Kruger A., Kummerfeld S.K.,
RA   Kurochkin I.V., Lareau L.F., Lazarevic D., Lipovich L., Liu J.,
RA   Liuni S., McWilliam S., Madan Babu M., Madera M., Marchionni L.,
RA   Matsuda H., Matsuzawa S., Miki H., Mignone F., Miyake S., Morris K.,
RA   Mottagui-Tabar S., Mulder N., Nakano N., Nakauchi H., Ng P.,
RA   Nilsson R., Nishiguchi S., Nishikawa S., Nori F., Ohara O.,
RA   Okazaki Y., Orlando V., Pang K.C., Pavan W.J., Pavesi G., Pesole G.,
RA   Petrovsky N., Piazza S., Reed J., Reid J.F., Ring B.Z., Ringwald M.,
RA   Rost B., Ruan Y., Salzberg S.L., Sandelin A., Schneider C.,
RA   Schonbach C., Sekiguchi K., Semple C.A., Seno S., Sessa L., Sheng Y.,
RA   Shibata Y., Shimada H., Shimada K., Silva D., Sinclair B.,
RA   Sperling S., Stupka E., Sugiura K., Sultana R., Takenaka Y., Taki K.,
RA   Tammoja K., Tan S.L., Tang S., Taylor M.S., Tegner J., Teichmann S.A.,
RA   Ueda H.R., van Nimwegen E., Verardo R., Wei C.L., Yagi K.,
RA   Yamanishi H., Zabarovsky E., Zhu S., Zimmer A., Hide W., Bult C.,
RA   Grimmond S.M., Teasdale R.D., Liu E.T., Brusic V., Quackenbush J.,
RA   Wahlestedt C., Mattick J.S., Hume D.A., Kai C., Sasaki D., Tomaru Y.,
RA   Fukuda S., Kanamori-Katayama M., Suzuki M., Aoki J., Arakawa T.,
RA   Iida J., Imamura K., Itoh M., Kato T., Kawaji H., Kawagashira N.,
RA   Kawashima T., Kojima M., Kondo S., Konno H., Nakano K., Ninomiya N.,
RA   Nishio T., Okada M., Plessy C., Shibata K., Shiraki T., Suzuki S.,
RA   Tagami M., Waki K., Watahiki A., Okamura-Oho Y., Suzuki H., Kawai J.,
RA   Hayashizaki Y.;
RT   "The transcriptional landscape of the mammalian genome.";
RL   Science 309:1559-1563(2005).
RN   [3]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   PubMed=16141073; DOI=10.1126/science.1112009;
RG   RIKEN Genome Exploration Research Group, and Genome Science Group
RG   (Genome Network Core Team) and the FANTOM Consortium;
RT   "Antisense Transcription in the Mammalian Transcriptome.";
RL   Science 309:1564-1566(2005).
RN   [4]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [5]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=21085660; PubMed=11217851; DOI=10.1038/35055500;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
RN   [6]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20499374; PubMed=11042159; DOI=10.1101/gr.145100;
RA   Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA   Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT   "Normalization and subtraction of cap-trapper-selected cDNAs to
RT   prepare full-length cDNA libraries for rapid discovery of new genes.";
```

```
RL   Genome Res. 10:1617-1630(2000).
RN   [7]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RX   MEDLINE=20530913; PubMed=11076861; DOI=10.1101/gr.152600;
RA   Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA   Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA   Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA   Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA   Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA   Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA   Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT   "RIKEN integrated sequence analysis (RISA) system-384-format
RT   sequencing pipeline with 384 multicapillary sequencer.";
RL   Genome Res. 10:1757-1771(2000).
RN   [8]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=C57BL/6J; TISSUE=Testis;
RA   Arakawa T., Carninci P., Fukuda S., Hashizume W., Hayashida K.,
RA   Hori F., Iida J., Imamura K., Imotani K., Itoh M., Kanagawa S.,
RA   Kawai J., Kojima M., Konno H., Murata M., Nakamura M., Ninomiya N.,
RA   Nishiyori H., Nomura K., Ohno M., Sakazume N., Sano H., Sasaki D.,
RA   Shibata K., Shiraki T., Tagami M., Tagami Y., Waki K., Watahiki A.,
RA   Muramatsu M., Hayashizaki Y.;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AK133099; BAE21510.1; -; mRNA.
DR   MGI; MGI:1922555; 4930564N15Rik.
DR   GO; GO:0003700; F:transcription factor activity; RCA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; RCA.
DR   InterPro; IPR002110; ANK.
DR   Pfam; PF00023; Ank; 3.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 3.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 2.
KW   ANK repeat; Hypothetical protein; Repeat.
SQ   SEQUENCE   497 AA;  56737 MW;  CA87BC48CC540354 CRC64;

  Query Match             8.0%;  Score 96;  DB 2;  Length 497;
  Best Local Similarity   21.1%;  Pred. No. 5.3;
  Matches   64;  Conservative   38;  Mismatches   89;  Indels  112;  Gaps   14;

Qy          4 QPLRHRSRCATPPRG-----------------------DFCGGTERA----------ID 29
              :| | ||: :||||                       ||| :|          :
Db          3 RPSRRRSKASTPPRSHTTPRRTGPSDSRRRPGTKEQPRPSVQGGTRQAEHDLKVSSPNSE 62

Qy         30 QASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQ----PER----------C 75
               : ::||  |  :|:   | |  | :|  :          ||:    |||
Db         63 SSQYSTSELWSHKVIANYSELFAASVGNVD---------WLRFCVNPERKEIIVDDKGFT 113

Qy         76 AVFQCAQ-----CHAVLADSVHLAWDLSRSLG-----AVVFSRVTNNVVLEAPFLVGIEG 125
              |:   ||      |  || :      ||  : |      |:     ::::    :|:
Db        114 AIHFAAQKCQLSCLKVLIEEYKYPVDLPTNKGQTPLHLVIHKNNKSDILPCIDYLLKKGA 173

Qy        126 SLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALR-------------------GHFCLSS 166
              ::   |||    ||  :    || | :  |  ::                      ::|
Db        174 AINSQTYN----GSTPL----HLASCNGLLGCIKLLVQSGANVHARDATGFKPINYCRLW 225

Qy        167 DKMVC------YLLKTKAIVNASEMDIQNVPLSEKIAELKEKI-VLTHNRLKSLMK---I 216
              :   |      : |    | | ||        ||:  ||||: :| :: |    |   |
Db        226 NHRTCARFLKDVMWKHDKKVLAQEM--------EKLRTLKEKLTILEYHYLVEYQKEHQI 277

Qy        217 LSE 219
              | |
Db        278 LRE 280


RESULT 13
Q4Q626_LEIMA
```

```
ID   Q4Q626_LEIMA   PRELIMINARY;   PRT;  1556 AA.
AC   Q4Q626;
DT   19-JUL-2005, integrated into UniProtKB/TrEMBL.
DT   19-JUL-2005, sequence version 1.
DT   07-FEB-2006, entry version 3.
DE   Hypothetical protein.
GN   ORFNames=LmjF31.2350;
OS   Leishmania major.
OC   Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX   NCBI_TaxID=5664;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RC   STRAIN=Friedlin;
RX   PubMed=16020728; DOI=10.1126/science.1112680;
RA   Ivens A.C., Peacock C.S., Worthey E.A., Murphy L., Aggarwal G.,
RA   Berriman M., Sisk E., Rajandream M.A., Adlem E., Aert R., Anupama A.,
RA   Apostolou Z., Attipoe P., Bason N., Bauser C., Beck A., Beverley S.M.,
RA   Bianchettin G., Borzym K., Bothe G., Bruschi C.V., Collins M.,
RA   Cadag E., Ciarloni L., Clayton C., Coulson R.M.R., Cronin A.,
RA   Cruz A.K., Davies R.M., De Gaudenzi J., Dobson D.E., Duesterhoeft A.,
RA   Fazelina G., Fosker N., Frasch A.C., Fraser A., Fuchs M., Gabel C.,
RA   Goble A., Goffeau A., Harris D., Hertz-Fowler C., Hilbert H., Horn D.,
RA   Huang Y., Klages S., Knights A., Kube M., Larke N., Litvin L.,
RA   Lord A., Louie T., Marra M., Masuy D., Matthews K., Michaeli S.,
RA   Mottram J.C., Mueller-Auer S., Munden H., Nelson S., Norbertczak H.,
RA   Oliver K., O'neil S., Pentony M., Pohl T.M., Price C., Purnelle B.,
RA   Quail M.A., Rabbinowitsch E., Reinhardt R., Rieger M., Rinta J.,
RA   Robben J., Robertson L., Ruiz J.C., Rutter S., Saunders D.,
RA   Schaefer M., Schein J., Schwartz D.C., Seeger K., Seyler A., Sharp S.,
RA   Shin H., Sivam D., Squares R., Squares S., Tosato V., Vogt C.,
RA   Volckaert G., Wambutt R., Warren T., Wedler H., Woodward J., Zhou S.,
RA   Zimmermann W., Smith D.F., Blackwell J.M., Stuart K.D., Barrell B.G.,
RA   Myler P.J.;
RT   "The genome of the kinetoplastid parasite, Leishmania major.";
RL   Science 309:436-442(2005).
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; CT005268; CAJ08424.1; -; Genomic_DNA.
KW   Hypothetical protein.
SQ   SEQUENCE   1556 AA;  170195 MW;  F6FF7CDDFE8DCA59 CRC64;

  Query Match             7.8%;  Score 94;  DB 2;  Length 1556;
  Best Local Similarity   24.3%;  Pred. No. 35;
  Matches   45;  Conservative   22;  Mismatches   50;  Indels   68;  Gaps    9;

Qy         22 GGTERAIDQASFTTSME------WDTQVVKGS----------SPLGPAGLGAEEPAAGPQ 65
              || |      | |:  :|      :|  | : |          | |  | | |: |||
Db        833 GGGELMQTDAGFSIEVESPIAFLYDDDVHRASQDVGQKRAREEPDGVGGAGEEDAAAG-- 890

Qy         66 LPSWLQPERCAVFQCAQCHAVLADSV-------HLAWDLSRSLGAVVFSRVTNNVVLEAP 118
                 | |: :  |     | |:|       |:   :   :     : :  :|:|:||
Db        891 -------ARAAIGE--QDFGFLNDTVAAPSTYSHMLCRVLELIRTARLNLILLDVLLQAP 941

Qy        119 FLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVC----YLL 174
              |          : |            | :|||           |||||:||  |    :||
Db        942 VL---------TRY----------VWYHLYK----------HFCLSADKASCVIGMWLL 971

Qy        175 KTKAI 179
              :  |:
Db        972 RNLAV 976


RESULT 14
Q2QSG3_ORYSA
ID   Q2QSG3_ORYSA   PRELIMINARY;   PRT;   702 AA.
AC   Q2QSG3;
DT   24-JAN-2006, integrated into UniProtKB/TrEMBL.
DT   24-JAN-2006, sequence version 1.
DT   07-FEB-2006, entry version 2.
DE   Retrotransposon protein, putative, unclassified.
GN   ORFNames=LOC_Os12g24210;
```

```
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Buell C.R., Wing R.A., McCombie W.A., Ouyang S.;
RL   Submitted (APR-2005) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; DP000011; ABA97657.1; -; Genomic_DNA.
SQ   SEQUENCE   702 AA;  76586 MW;  5E958D36E696F434 CRC64;

  Query Match             7.8%;  Score 93.5;  DB 2;  Length 702;
  Best Local Similarity   26.2%;  Pred. No. 14;
  Matches   67;  Conservative   29;  Mismatches   69;  Indels   91;  Gaps   18;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLP----SW--LQPERC 75
              |||  |   |: |     |  :  ||    |:|    :||||   |    :|  || |
Db        335 GGTSSASPAAAST-----DVALATGSREATPSG-PVSDPAAGRGPPAAVLTWEELQVEMR 388

Qy         76 AVFQC----------AQCHAVLADSVHLAW----------DLS--RSLGAV-------- 104
               : :           |:  | ||:: |  |          ||:  | | ||
Db        389 RLLKAGARSIGREIAEARSEAALANT-HADWLVRELAEAREDLTKMRELVAVNERQRKGL 447

Qy        105 --VFSRVTNNVVLEAPFLVGIEGSLKGSTYNLL--FCGSCGI-------PVGFHLYSTHA 153
                 | : ||       |  | |||:  ||  |    | |||       |  | | |: |
Db        448 EDRMSELGNN-------LSEIRGSLR-VTYTGLHQLAGECGIKSTIPVNPDEFSLNSSLA 499

Qy        154 ALAALRG-----HFCLSSDKM-------VCYLLKTKAIVNAS--EMDIQNV--------- 190
               || : |     |    :::        |::|   | |  |  |:|:: :
Db        500 ELATVMGEIPSKHAARIAEETSNGIYTGACHVL---ACVKLSRPELDLREILDQGAASDT 556

Qy        191 --PLSEKIAELKEKIV 204
                 : ||: :| | ::
Db        557 REEVMEKVGDLGESVL 572


RESULT 15
Q8S611_ORYSA
ID   Q8S611_ORYSA   PRELIMINARY;   PRT;   361 AA.
AC   Q8S611;
DT   01-JUN-2002, integrated into UniProtKB/TrEMBL.
DT   01-JUN-2002, sequence version 1.
DT   21-FEB-2006, entry version 13.
DE   Putative gypsy-type retrotransposon protein.
GN   Name=OSJNBa0096E22.5;
OS   Oryza sativa (japonica cultivar-group).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; BEP clade;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=39947;
RN   [1]
RP   NUCLEOTIDE SEQUENCE.
RA   Wing R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA   Saski C., Henry D., Oates R., Simmons J.;
RT   "Rice Genomic Sequence.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   ---------------------------------------------------------------------
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   ---------------------------------------------------------------------
DR   EMBL; AC099400; AAL91599.1; -; Genomic_DNA.
DR   Gramene; Q8S611; -.
SQ   SEQUENCE   361 AA;  38805 MW;  F883BB3E8B8FFF45 CRC64;

  Query Match             7.7%;  Score 92;  DB 2;  Length 361;
  Best Local Similarity   25.3%;  Pred. No. 8.4;
  Matches   61;  Conservative   28;  Mismatches   92;  Indels   60;  Gaps   14;

Qy          2 AAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPA 61
              |  ||  |          |  |||  |    : |      |  || ||    |:|  | :|
Db         49 APTPLPFRRAVRAKAAQDGSGGTSSASPAVAST-----DVVVVPGSREATPSG-PASDPV 102

Qy         62 AGPQLP----SW--LQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVL 115
              ||  |    ||  || |   : :      |:  :  |  |   :     | : ||
Db        103 AGRGSPAAVLSWEELQVEMGRLLEAGA--RVIGREIAEARGLEHRM-----SELGNN--- 152

Qy        116 EAPFLVGIEGSLKGSTYNLL--FCGSCGI-------PVGFHLYSTHAALAAL-----RGH 161
                | | |||:  || |    | |||       |  | | |: | |||       |
Db        153 ----LSEIRGSLR-VTYTGLHQLAGKCGIKSTIPANPDEFSLTSSLAELAAAMEEIPSKH 207

Qy        162 FCLSSDKM-------VCYLLKTKAIVNASEMDIQNV----------PLSEKIAELKEKI 203
                ::|         |::|    : :  |:|:: :           : |:: :| : :
Db        208 AARIGEEMSNRIYIGACHILACVRLAH-PELDLREILDQGEASDARKDVMEEVGDLGKSV 266

Qy        204 V 204
                :
Db        267 L 267


Search completed: May 18, 2006, 15:40:51
Job time : 302 secs
```

GenCore version 5.1.8

OM protein - protein search, using sw model

Run on: May 18, 2006, 15:41:07 ; Search time 50 Seconds
(without alignments)
400.890 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /EMC_Celerra_SIDS3/ptodata/2/iaa/5_COMB.pep:*
2: /EMC_Celerra_SIDS3/ptodata/2/iaa/6_COMB.pep:*
3: /EMC_Celerra_SIDS3/ptodata/2/iaa/7_COMB.pep:*
4: /EMC_Celerra_SIDS3/ptodata/2/iaa/H_COMB.pep:*
5: /EMC_Celerra_SIDS3/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /EMC_Celerra_SIDS3/ptodata/2/iaa/RE_COMB.pep:*
7: /EMC_Celerra_SIDS3/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	86	7.2	158	2	US-09-134-000C-6596	Sequence 6596, Ap
2	83	6.9	457	7	5268463-7	Patent No. 5268463
3	81.5	6.8	456	7	5432081-7	Patent No. 5432081
4	81.5	6.8	457	1	US-08-882-704A-6	Sequence 6, Appli
5	81.5	6.8	457	2	US-09-151-957-6	Sequence 6, Appli
6	81.5	6.8	457	2	US-10-195-518-6	Sequence 6, Appli
7	80.5	6.7	1711	1	US-08-342-930-2	Sequence 2, Appli
8	78.5	6.6	325	1	US-09-018-576-3	Sequence 3, Appli
9	78.5	6.6	325	1	US-09-018-576-12	Sequence 12, Appl
10	78.5	6.6	325	2	US-09-248-137-3	Sequence 3, Appli
11	78.5	6.6	325	2	US-09-248-137-12	Sequence 12, Appl
12	78.5	6.6	452	2	US-09-771-161A-237	Sequence 237, App
13	78	6.5	190	2	US-09-252-991A-19049	Sequence 19049, A
14	77	6.4	607	2	US-09-537-682-1	Sequence 1, Appli
15	77	6.4	719	1	US-08-520-933-3	Sequence 3, Appli
16	77	6.4	719	2	US-09-285-040-3	Sequence 3, Appli
17	77	6.4	738	7	5262177-2	Patent No. 5262177
18	77	6.4	776	2	US-09-949-016-10829	Sequence 10829, A
19	76.5	6.4	282	2	US-09-845-713A-2	Sequence 2, Appli
20	76	6.3	351	2	US-09-902-540-13214	Sequence 13214, A
21	76	6.3	1001	2	US-10-104-047-2283	Sequence 2283, Ap
22	75.5	6.3	310	1	US-08-129-456A-36	Sequence 36, Appl
23	75.5	6.3	310	1	US-08-705-868-3	Sequence 3, Appli
24	75.5	6.3	310	2	US-09-123-615-3	Sequence 3, Appli
25	75.5	6.3	310	2	US-08-360-821B-35	Sequence 35, Appl
26	75.5	6.3	310	2	US-09-851-026-35	Sequence 35, Appl
27	75.5	6.3	1449	2	US-08-840-062-6	Sequence 6, Appli
28	75.5	6.3	1723	2	US-09-194-612A-31	Sequence 31, Appl
29	75.5	6.3	4545	1	US-08-804-227C-14	Sequence 14, Appl
30	75.5	6.3	4550	1	US-08-804-227C-8	Sequence 8, Appli
31	75.5	6.3	4550	1	US-08-804-198-2	Sequence 2, Appli
32	75	6.3	78	2	US-09-270-767-43496	Sequence 43496, A
33	75	6.3	458	2	US-09-800-170-3	Sequence 3, Appli
34	75	6.3	717	7	5262177-5	Patent No. 5262177
35	75	6.3	784	2	US-09-724-586A-1	Sequence 1, Appli
36	75	6.3	784	2	US-09-600-823-1	Sequence 1, Appli
37	75	6.3	784	2	US-09-724-666-1	Sequence 1, Appli
38	75	6.3	1088	2	US-09-920-804-2	Sequence 2, Appli
39	75	6.3	1088	2	US-09-920-804-4	Sequence 4, Appli
40	75	6.3	1129	2	US-09-734-674-2	Sequence 2, Appli
41	75	6.3	1129	2	US-10-274-990-2	Sequence 2, Appli
42	75	6.3	2227	2	US-08-475-886-4	Sequence 4, Appli
43	75	6.3	2227	2	US-09-653-499-4	Sequence 4, Appli
44	75	6.3	2227	2	US-10-135-988-4	Sequence 4, Appli
45	74.5	6.2	211	2	US-09-949-016-11409	Sequence 11409, A

ALIGNMENTS

RESULT 1
US-09-134-000C-6596
; Sequence 6596, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6596
; LENGTH: 158
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6596

Query Match 7.2%; Score 86; DB 2; Length 158;
Best Local Similarity 26.5%; Pred. No. 0.19;
Matches 40; Conservative 27; Mismatches 54; Indels 30; Gaps 9;

```
Qy          87 LADSVHLAWDLSRSLGAVV---FSRVTNNVVLEA--PFLVGIEGSLKG--------STYN 133
               | : ::|| :  | |::     |: | |:: :|   |:  ||  |         | |
Db          17 LTEYIYLASEQPLSTGSITLNGLSKPTGNLLNKAFFFFINDFEGDPPGKFSYSNYLKTLN 76

Qy         134 LLFCGSCGIPVGFHLYSTHAALAAL---RGHFCLSSDKMVCYLLKTKAIVNASEMDIQNV 190
                 |  | :|       | :|:| |   | ||     |  |    ||  : ::    :|
Db          77 YQVCSMCSLPSDNQNIETPSAIALLGYIRNHF-----KNTC---TTKLYILSALNGFENK 128

Qy         191 PLSE----KIAELKEKIVLTHNRLKSLMKIL 217
               |:|:    |:::|  |  | :  :| |:||:
Db         129 PISKTVTRKLSDLSYK-DLYYQEMK-LLKII 157
```

RESULT 2
5268463-7
;Patent No. 5268463
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: PLAANT PROMOTER a-GLUCURONIDASE GENE
;CONSTRUCT
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/447,976
; FILING DATE: 08-DEC-1989
; PRIOR APPLICATION DATA:

```
;      APPLICATION NUMBER: 119,102
;      FILING DATE: 10-NOV-1987
;      APPLICATION NUMBER: 264,586
;      FILING DATE: 31-OCT-1988
;SEQ ID NO:7:
;      LENGTH: 457
5268463-7

  Query Match              6.9%;  Score 83;  DB 7;  Length 457;
  Best Local Similarity   25.2%;  Pred. No. 2.2;
  Matches   37;  Conservative   30;  Mismatches   54;  Indels   26;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        316 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 373

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        374 QAIGGS--SIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKETVVEIDNRKKVQQQLISDIT 456


RESULT 3
5432081-7
;Patent No. 5432081
;    APPLICANT: JEFFERSON, RICHARD A.
;    TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E.COLI
;GLUCORONIDE PERMEASE GENE
;    NUMBER OF SEQUENCES: 10
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/138,546
;      FILING DATE: 15-OCT-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 447,976
;      FILING DATE: 08-DEC-1989
;      APPLICATION NUMBER: 264,586
;      FILING DATE: 31-OCT-1988
;      APPLICATION NUMBER: 119,102
;      FILING DATE: 10-NOV-1987
;SEQ ID NO:7:
;      LENGTH: 456
5432081-7

  Query Match              6.8%;  Score 81.5;  DB 7;  Length 456;
  Best Local Similarity   25.2%;  Pred. No. 3.2;
  Matches   37;  Conservative   30;  Mismatches   53;  Indels   27;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        316 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 373

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        374 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 430

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        431 K--KFKEIVVEIDNRKKVQQQLISDIT 455


RESULT 4
US-08-882-704A-6
; Sequence 6, Application US/08882704A
; Patent No. 5879906
;  GENERAL INFORMATION:
;    APPLICANT:  Jefferson, Richard A.
;    APPLICANT:  Wilson, Katherine J.
;    APPLICANT:  Leader, Michael
;    TITLE OF INVENTION:  GLUCURONIDE REPRESSORS AND USES THEREOF
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/882,704A
;      FILING DATE:  25-JUN-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5879906tenburg Ph.D., Carol
;      REGISTRATION NUMBER:  39,317
;      REFERENCE/DOCKET NUMBER:  190106.404
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  457 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-882-704A-6

  Query Match              6.8%;  Score 81.5;  DB 1;  Length 457;
  Best Local Similarity   25.2%;  Pred. No. 3.2;
  Matches   37;  Conservative   30;  Mismatches   53;  Indels   27;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
                :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456


RESULT 5
US-09-151-957-6
; Sequence 6, Application US/09151957
; Patent No. 6429292
;  GENERAL INFORMATION:
;       APPLICANT: Jefferson, Richard A.
;                  Wilson, Katherine J.
;                  Leader, Michael
;       TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF
;       NUMBER OF SEQUENCES: 19
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: SEED and BERRY LLP
;            STREET: 6300 Columbia Center, 701 Fifth Avenue
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/151,957
;            FILING DATE: 11-Sep-1998
```

```
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US 08/882,704
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: No. 6429292tenburg Ph.D., Carol
;            REGISTRATION NUMBER: 39,317
;            REFERENCE/DOCKET NUMBER: 190106.404
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 457 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-151-957-6

  Query Match             6.8%;  Score 81.5;  DB 2;  Length 457;
  Best Local Similarity   25.2%;  Pred. No. 3.2;
  Matches   37;  Conservative   30;  Mismatches   53;  Indels   27;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
              :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456


RESULT 6
US-10-195-518-6
; Sequence 6, Application US/10195518
; Patent No. 6998229
;   GENERAL INFORMATION:
;       APPLICANT: Jefferson, Richard A.
;                  Wilson, Katherine J.
;                  Leader, Michael
;       TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF
;       NUMBER OF SEQUENCES: 19
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: SEED and BERRY LLP
;            STREET: 6300 Columbia Center, 701 Fifth Avenue
;            CITY: Seattle
;            STATE: Washington
;            COUNTRY: USA
;            ZIP: 98104-7092
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/195,518
;            FILING DATE: 16-Jul-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/151,957
;            FILING DATE: 11-Sep-1998
;            APPLICATION NUMBER: US 08/882,704
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: No. 6998229tenburg Ph.D., Carol
;            REGISTRATION NUMBER: 39,317
;            REFERENCE/DOCKET NUMBER: 190106.404
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (206) 622-4900
;            TELEFAX: (206) 682-6031
;   INFORMATION FOR SEQ ID NO: 6:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 457 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-195-518-6

  Query Match             6.8%;  Score 81.5;  DB 2;  Length 457;
  Best Local Similarity   25.2%;  Pred. No. 3.2;
  Matches   37;  Conservative   30;  Mismatches   53;  Indels   27;  Gaps    8;

Qy         95 WDLSRSLGAVVFSRVTNNV------VLEA------PFLVGIEGSLKGSTYNLL-FCGSCG 141
              | |  :| |:  : :   |       |||       :| |:   ::| ||:|  |   ||
Db        317 WSLPVALVALAIASIGQGVTMTVMWALEADTVEYGEYLTGVR--IEGLTYSLFSFTRKCG 374

Qy        142 IPVGFHLYSTHAALAALRGHFC--LSSDKMVCYLLKTKAIVNASEMDIQNV-----PLSE 194
              :|    |  | :  | |:    : : :::  :  : |:|    | :  |     ||::
Db        375 QAIG---GSIPAFILGLSGYIANQVQTPEVIMGIRTSIALVPCGFMLLAFVIIWFYPLTD 431

Qy        195 KIAELKEKIVLTHNRLKSLMKILSEVT 221
              |  : || :|   || |   :::|::|
Db        432 K--KFKEIVVEIDNRKKVQQQLISDIT 456


RESULT 7
US-08-342-930-2
; Sequence 2, Application US/08342930
; Patent No. 5821084
;  GENERAL INFORMATION:
;    APPLICANT:  OLMSTED, ELIZABETH A.
;    APPLICANT:  MAURO, LAURA J.
;    APPLICANT:  DAVIS, ALAN R.
;    APPLICANT:  DIXON, JACK E.
;    TITLE OF INVENTION:  OSTEOBLAST-TESTICULAR PROTEIN TYROSINE
;    TITLE OF INVENTION:  PHOSPHATASE
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 Page Mill Road
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/342,930
;      FILING DATE:  21-NOV-1994
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  KONSKI, ANTOINETTE F.
;      REGISTRATION NUMBER:  34,202
;      REFERENCE/DOCKET NUMBER:  20344-20975.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 813-5600
;      TELEFAX:  (415) 494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1711 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-342-930-2

  Query Match             6.7%;  Score 80.5;  DB 1;  Length 1711;
```

```
  Best Local Similarity   29.6%;  Pred. No. 31;
  Matches   21;  Conservative    8;  Mismatches   17;  Indels   25;  Gaps    1;

Qy         36 SMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAW 95
              ||:  : ||  : | || | ||                              ||| |:|: ||:|
Db        753 SMQAGSAVVNLAWPSGPLGQGA------------------------CHAQLSDAGHLSW 787

Qy         96 DLSRSLGAVVF 106
              :   ||  :|
Db        788 EQPLKLGQELF 798


RESULT 8
US-09-018-576-3
; Sequence 3, Application US/09018576
; Patent No. 5968800
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,576
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-018-576-3

  Query Match             6.6%;  Score 78.5;  DB 1;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 4.1;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
              |   :|: :       || | | |     : | | |                 |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
              |: |:||     |   | |||                            ||:|     |
Db        102 AQVKSYLQMLLKGV---AFCHA------------------------NNIVHRDLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
              |:   | || : : |    |  | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                : :||:|:     |    ::  :|   ||:  ||:: :          | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
              |
Db        237 S 237


RESULT 9
US-09-018-576-12
; Sequence 12, Application US/09018576
; Patent No. 5968800
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/018,576
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-018-576-12

  Query Match             6.6%;  Score 78.5;  DB 1;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 4.1;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
              |   :|: :       || | | |     : | | |                 |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
              |: |:||     |   | |||                            ||:|     |
Db        102 AQVKSYLQMLLKGV---AFCHA------------------------NNIVHRNLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
              |:   | || : : |    |  | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                : :||:|:     |    ::  :|   ||:  ||:: :          | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
              |
Db        237 S 237


RESULT 10
US-09-248-137-3
; Sequence 3, Application US/09248137
; Patent No. 6030788
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/248,137
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/018,576
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-248-137-3

  Query Match             6.6%;  Score 78.5;  DB 2;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 4.1;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
                |   :|: :      || | | |     : | | |                 |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
               |: |:||     |   | |||                          ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA-----------------------NNIVHRDLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD----------- 167
               |:   | || : : |     |   | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                 : :||:|:     |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
               |
Db        237 S 237


RESULT 11
US-09-248-137-12
; Sequence 12, Application US/09248137
; Patent No. 6030788
;  GENERAL INFORMATION:
;    APPLICANT:  Gerhold, David L.
;    TITLE OF INVENTION:  CYCLIN-DEPENDENT PROTEIN KINASE
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Merck & Co., Inc.
;      STREET:  P.O. Box 2000, RY60-30
;      CITY:  Rahway
;      STATE:  NJ
;      COUNTRY:  US
;      ZIP:  07065-0907
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/248,137
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  09/018,576
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Hand, J. Mark
;      REGISTRATION NUMBER:  36,545
;      REFERENCE/DOCKET NUMBER:  19885Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  732/594-3905
;      TELEFAX:  732/594-4720
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  325 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-248-137-12

  Query Match             6.6%;  Score 78.5;  DB 2;  Length 325;
  Best Local Similarity   23.2%;  Pred. No. 4.1;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
                |   :|: :      || | | |     : | | |                 |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
               |: |:||     |   | |||                          ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA-----------------------NNIVHRNLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD----------- 167
               |:   | || : : |     |   | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                 : :||:|:     |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
               |
Db        237 S 237


RESULT 12
US-09-771-161A-237
; Sequence 237, Application US/09771161A
; Patent No. 6936450
; GENERAL INFORMATION:
;  APPLICANT: LEVINE, et al.
;  TITLE OF INVENTION: VARIANTS Of PROTEIN KINASES
;  FILE REFERENCE: 802620-2005.1
;  CURRENT APPLICATION NUMBER: US/09/771,161A
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: 09/724,676
;  PRIOR FILING DATE: 2000-11-28
;  PRIOR APPLICATION NUMBER: 136776
;  PRIOR FILING DATE: 2000-06-15
```

```
;  PRIOR APPLICATION NUMBER: 135619
;  PRIOR FILING DATE: 2000-04-12
;  NUMBER OF SEQ ID NOS: 273
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 237
;   LENGTH: 452
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-771-161A-237

  Query Match             6.6%;  Score 78.5;  DB 2;  Length 452;
  Best Local Similarity   23.2%;  Pred. No. 6.9;
  Matches   56;  Conservative   24;  Mismatches   78;  Indels   83;  Gaps   12;

Qy         22 GGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG-----------------AEEPAAG 63
              |   :|: :      || | | |     : | | |                 |: | |
Db         43 GFPNQALREIKALQEME-DNQYVVQLKAVFPHGGGFVLAFEFMLSDLAEVVRHAQRPLAQ 101

Qy         64 PQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVV----LEAPF 119
               |: |:||      |   | |||                          ||:|      |
Db        102 AQVKSYLQMLLKGV---AFCHA-----------------------NNIVHRDLKPANL 133

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSD------------ 167
              |:   | || : : |    |  | |  || ||    | |   |:  :
Db        134 LISASGQLKIADFGLARVFS---PDGSRLY-THQ--VATRSVGCIMGELLNGSPLFPGKN 187

Qy        168 --KMVCYLLKTKAIVNASEM-DIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQ 224
                : :||:|:     |     ::  :|   ||:  ||:: :           | || ||
Db        188 DIEQLCYVLRILGTPNPQVWPELTELPDYNKIS-FKEQVPMP----------LEEVLPDV 236

Qy        225 S 225
              |
Db        237 S 237


RESULT 13
US-09-252-991A-19049
; Sequence 19049, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 19049
;   LENGTH: 190
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19049

  Query Match             6.5%;  Score 78;  DB 2;  Length 190;
  Best Local Similarity   32.9%;  Pred. No. 2.1;
  Matches   24;  Conservative    6;  Mismatches   29;  Indels   14;  Gaps    3;

Qy          4 QPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAG 63
              :| ||  ||  |  | : |   || |:||          | : : |  ||  |
Db         94 RPCRHTRRCVRPAAGGYPG---RAGDRAS----------VARSTGPGLPAAPGRPRGHC- 139

Qy         64 PQLPSWLQPERCA 76
              |  | | :  |||
Db        140 PAPPRWPRRRRCA 152


RESULT 14
US-09-537-682-1
; Sequence 1, Application US/09537682
; Patent No. 6303357
; GENERAL INFORMATION:
;  APPLICANT: TAKEUCHI, Kenichi
;  APPLICANT: KOIDE, Yoshinao
;  APPLICANT: NAKANISHI, Yuji
;  APPLICANT: SUZUKI, Satoru
;  TITLE OF INVENTION: L-ALPHA-GLYCEROPHOSPHATE OXIDASE GENE, RECOMBINATNT
;  TITLE OF INVENTION: DNA, AND METHOD FOR PRODUCING MODIFIED
;  TITLE OF INVENTION: L-ALPHA-GLYCEROPHOSPHATE OXIDASE GENE
;  FILE REFERENCE: A20-121814C/KI
;  CURRENT APPLICATION NUMBER: US/09/537,682
;  CURRENT FILING DATE: 2000-03-29
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 607
;   TYPE: PRT
;   ORGANISM: Enterococcus faecium No. 6303357 7044
US-09-537-682-1

  Query Match             6.4%;  Score 77;  DB 2;  Length 607;
  Best Local Similarity   23.2%;  Pred. No. 16;
  Matches   49;  Conservative   30;  Mismatches   76;  Indels   56;  Gaps   10;

Qy         19 DFCGGTERAIDQASFTTSM---------EWDTQVVKGSSPLGP----AGLGAEEPAAGPQ 65
              ||  ||      :| :| :           :| :||  :  :|      |:    |  |
Db         51 DFAEGT------SSRSTKLVHGGIRYLKTFDVEVV--ADTVGERAVVQGIAPHIPKPDPM 102

Qy         66 LPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTN------NVVLEAPF 119
              |    : |    |            || :| ||   | |  ::  |       |:   ||
Db        103 LLPIYEDEGATTFNMF--------SVKVAMDLYDKLANVTGTKYENYTLTPEEVLEREPF 154

Qy        120 LVGIEGSLKGSTYNLLFCGSCGIPVGF-----HLYSTHAALAALRGHFCLSSDKMVCYLL 174
              |   :  |||:          |: : |      |   :   ||  | : :|  | | :|
Db        155 LK--KEGLKGA----------GVYLDFRNNDARLVIDNIKKAAEDGAYLVSKMKAVGFLY 202

Qy        175 KTKAIVNASEMDIQNVPLSEKIAELKEKIVL 205
              :   ||      |:    |:::: |:| |:|:
Db        203 EGDQIVGVKARDL----LTDEVIEIKSKLVI 229


RESULT 15
US-08-520-933-3
; Sequence 3, Application US/08520933
; Patent No. 5981194
;  GENERAL INFORMATION:
;    APPLICANT:  Jefferies, Wilfred A.
;    APPLICANT:  McGeer, Patrick L.
;    APPLICANT:  Rothenberger, Sylvia
;    APPLICANT:  Food, Michael R.
;    APPLICANT:  Yamada, Tatsuo
;    APPLICANT:  Kennard, Malcolm
;    TITLE OF INVENTION:  Use of p97 and Iron Binding Proteins
;    TITLE OF INVENTION:  as Diagnostic and Therapeutic Agents
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bereskin & Parr
;      STREET:  40 King Street West
;      CITY:  Toronto
;      STATE:  Ontario
;      COUNTRY:  Canada
;      ZIP:  M5H 3Y2
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/520,933
;      FILING DATE:  August 31, 1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Shona S. McDiarmid
```

; REGISTRATION NUMBER: 38,798
; REFERENCE/DOCKET NUMBER: 7685-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 416-364-7311
; TELEFAX: 416-361-1398
; TELEX: 06-23115
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 719 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-520-933-3

Query Match 6.4%; Score 77; DB 1; Length 719;
Best Local Similarity 23.7%; Pred. No. 21;
Matches 59; Conservative 25; Mismatches 87; Indels 78; Gaps 15;

```
Qy          17 RGDFCGGTERAIDQASFTTSMEW----------DTQVVKGSSPLGPAGLGAEEPAAGP 64
               |||  |  |   |::      ::             |   || |: |         |   |
Db         171 RGDSSG--EGVCDKSPLERYYDYSGAFRCLAEGAGDVAFVKHSTVL--------ENTDGK 220

Qy          65 QLPSWLQP-------------ERCAVFQCAQCH--AVLADSVHLAWDLSRSLGAVVFSRV 109
                ||||  |               |  | :  |||   | | :| :  |     | ::| |:
Db         221 TLPSWGQALLSQDFELLCRDGSRADVTEWRQCHLARVPAHAVVVRADTD---GGLIF-RL 276

Qy         110 TNNVVLEAPFLVGIEGS----LKGSTY---NLLFCGSCG--IPVGFHLYST---HAALAA 157
                |    |   |   |||         |   :|||  |    :|:    |     |  | |
Db         277 LN----EGQRLFSHEGSSFQMFSSEAYGQKDLLFKDSTSELVPIATQTYEAWLGHEYLHA 332

Qy         158 LRGHFCLSSDKMVCYL----LKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLKSL 213
               ::|  |   :::  ||    | |  |    :|                :     |||
Db         333 MKGLLC-DPNRLPPYLRWCVLSTPEIQKCGDM---------------AVAFRRQRLKPE 375

Qy         214 MKILSEVTP 222
               :: :|  :|
Db         376 IQCVSAKSP 384
```

Search completed: May 18, 2006, 15:42:30
Job time : 51 secs

OM protein - protein search, using sw model

Run on: May 18, 2006, 15:52:57 ; Search time 182 Seconds
(without alignments)
582.836 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_Main:*
1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1198	100.0	229	3	US-09-942-052-728	Sequence 728, App
2	1198	100.0	229	3	US-09-942-052-729	Sequence 729, App
3	1198	100.0	229	3	US-09-942-052-730	Sequence 730, App
4	1198	100.0	229	3	US-09-942-052-731	Sequence 731, App
5	138.5	11.6	164	3	US-09-942-052-707	Sequence 707, App
6	138	11.5	29	3	US-09-942-052-704	Sequence 704, App
7	113	9.4	233	3	US-09-890-688-82	Sequence 82, Appl
8	113	9.4	233	4	US-10-408-765A-969	Sequence 969, App
9	95.5	8.0	708	4	US-10-437-963-195427	Sequence 195427,
10	92	7.7	361	4	US-10-437-963-154548	Sequence 154548,
11	92	7.7	695	4	US-10-369-493-18389	Sequence 18389, A
12	89	7.4	664	3	US-09-780-525-2	Sequence 2, Appli
13	88.5	7.4	435	4	US-10-437-963-186569	Sequence 186569,
14	87.5	7.3	848	4	US-10-437-963-155606	Sequence 155606,
15	86.5	7.2	669	4	US-10-282-122A-51802	Sequence 51802, A
16	86	7.2	584	6	US-11-096-568A-12723	Sequence 12723, A
17	86	7.2	619	6	US-11-096-568A-12722	Sequence 12722, A
18	86	7.2	644	6	US-11-096-568A-12721	Sequence 12721, A
19	85	7.1	337	4	US-10-156-761-10954	Sequence 10954, A
20	84.5	7.1	485	4	US-10-363-829-358	Sequence 358, App
21	84	7.0	295	4	US-10-424-599-223506	Sequence 223506,
22	84	7.0	1902	4	US-10-437-963-139559	Sequence 139559,
23	83.5	7.0	888	4	US-10-437-963-147897	Sequence 147897,
24	83	6.9	663	4	US-10-282-122A-51673	Sequence 51673, A
25	82.5	6.9	124	4	US-10-425-115-216789	Sequence 216789,
26	82.5	6.9	278	4	US-10-425-114-42310	Sequence 42310, A
27	82.5	6.9	325	4	US-10-663-896-2	Sequence 2, Appli
28	82.5	6.9	493	5	US-10-213-974-36	Sequence 36, Appl
29	82.5	6.9	502	4	US-10-424-599-226730	Sequence 226730,
30	82.5	6.9	1043	4	US-10-369-493-9793	Sequence 9793, Ap
31	81.5	6.8	278	4	US-10-425-114-66220	Sequence 66220, A
32	81.5	6.8	457	4	US-10-195-518-6	Sequence 6, Appli
33	81.5	6.8	701	5	US-10-450-763-31748	Sequence 31748, A
34	81.5	6.8	1969	4	US-10-437-963-115342	Sequence 115342,
35	81.5	6.8	3067	5	US-10-631-467-1618	Sequence 1618, Ap
36	81	6.8	229	4	US-10-425-115-281194	Sequence 281194,
37	81	6.8	859	4	US-10-210-281-60	Sequence 60, Appl
38	80.5	6.7	282	5	US-10-450-763-51058	Sequence 51058, A
39	80.5	6.7	1711	5	US-10-029-345A-38	Sequence 38, Appl
40	80.5	6.7	1711	6	US-11-143-984A-38	Sequence 38, Appl
41	80	6.7	250	6	US-11-079-463-9254	Sequence 9254, Ap
42	80	6.7	305	4	US-10-183-116-59	Sequence 59, Appl
43	80	6.7	305	5	US-10-957-135-59	Sequence 59, Appl
44	80	6.7	305	6	US-11-083-611-59	Sequence 59, Appl
45	80	6.7	544	4	US-10-424-599-158084	Sequence 158084,

ALIGNMENTS

RESULT 1
US-09-942-052-728
; Sequence 728, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
; APPLICANT: Raitano, Arthur B.
; APPLICANT: Faris, Mary
; APPLICANT: Hubert, Rene S.
; APPLICANT: Afar, Daniel
; APPLICANT: Ge, Wangmao
; APPLICANT: Challita-Eid, Pia M.
; TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
; TITLE OF INVENTION: USEFUL IN TREATMENT AND DETECTION OF CANCER
; FILE REFERENCE: 51158-20028.00
; CURRENT APPLICATION NUMBER: US/09/942,052
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: 60/228,432
; PRIOR FILING DATE: 2000-08-28
; NUMBER OF SEQ ID NOS: 744
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 728
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: 85P1B3/OIP5
; OTHER INFORMATION: clone A protein
US-09-942-052-728

Query Match 100.0%; Score 1198; DB 3; Length 229;
Best Local Similarity 100.0%; Pred. No. 1.1e-113;
Matches 229; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

```
RESULT 2
US-09-942-052-729
; Sequence 729, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 729
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3/OIP5
;   OTHER INFORMATION: protein sequence
US-09-942-052-729

  Query Match            100.0%;  Score 1198;  DB 3;  Length 229;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-113;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229


RESULT 3
US-09-942-052-730
; Sequence 730, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 730
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3 protein
;   OTHER INFORMATION: sequence
US-09-942-052-730

  Query Match            100.0%;  Score 1198;  DB 3;  Length 229;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-113;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229


RESULT 4
US-09-942-052-731
; Sequence 731, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 731
;   LENGTH: 229
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-942-052-731

  Query Match            100.0%;  Score 1198;  DB 3;  Length 229;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-113;
  Matches  229;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP 60

Qy         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 AAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVLEAPFL 120

Qy        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 VGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIV 180

Qy        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 NASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKPEN 229
```

```
RESULT 5
US-09-942-052-707
; Sequence 707, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 707
;   LENGTH: 164
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: Splice variant 1
;   OTHER INFORMATION: open reading frame 3 peptide sequence
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (44)
;   OTHER INFORMATION: Unknown amino acid or stop
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (50)..(51)
;   OTHER INFORMATION: Unknown amino acid or stop
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (67)
;   OTHER INFORMATION: Unknown amino acid or stop
;   FEATURE:
;   NAME/KEY: MOD_RES
;   LOCATION: (93)
;   OTHER INFORMATION: Unknown amino acid or stop
US-09-942-052-707

  Query Match             11.6%;  Score 138.5;  DB 3;  Length 164;
  Best Local Similarity   82.1%;  Pred. No. 1.3e-05;
  Matches   32;  Conservative    1;  Mismatches    1;  Indels    5;  Gaps    1;

Qy        162 FCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              | |||    |||||||||||||||||||||||||||:|
Db          2 FFLSS-----YLLKTKAIVNASEMDIQNVPLSEKIAEVK 35


RESULT 6
US-09-942-052-704
; Sequence 704, Application US/09942052
; Publication No. US20030170626A1
; GENERAL INFORMATION:
;  APPLICANT: Raitano, Arthur B.
;  APPLICANT:  Faris, Mary
;  APPLICANT:  Hubert, Rene S.
;  APPLICANT:  Afar, Daniel
;  APPLICANT:  Ge, Wangmao
;  APPLICANT:  Challita-Eid, Pia M.
;  TITLE OF INVENTION: NUCLEIC ACID AND CORRESPONDING PROTEIN ENTITLED 85P1B3
;  TITLE OF INVENTION:  USEFUL IN TREATMENT AND DETECTION OF CANCER
;  FILE REFERENCE: 51158-20028.00
;  CURRENT APPLICATION NUMBER: US/09/942,052
;  CURRENT FILING DATE:  2001-08-28
;  PRIOR APPLICATION NUMBER: 60/228,432
;  PRIOR FILING DATE: 2000-08-28
;  NUMBER OF SEQ ID NOS: 744
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 704
;   LENGTH: 29
;   TYPE: PRT
;   ORGANISM: Unknown Organism
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: 85P1B3 peptide
;   OTHER INFORMATION: sequence
US-09-942-052-704

  Query Match             11.5%;  Score 138;  DB 3;  Length 29;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-06;
  Matches   29;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        172 YLLKTKAIVNASEMDIQNVPLSEKIAELK 200
              |||||||||||||||||||||||||||||
Db          1 YLLKTKAIVNASEMDIQNVPLSEKIAELK 29


RESULT 7
US-09-890-688-82
; Sequence 82, Application US/09890688
; Publication No. US20030144475A1
; GENERAL INFORMATION:
;  APPLICANT: Seishi KATO
;  APPLICANT:  Chikashi EGUCHI
;  APPLICANT:  Mihoro SAEKI
;  TITLE OF INVENTION: Human Proteins and cDNAs thereof
;  FILE REFERENCE: 2001-1102A/WMC/00653
;  CURRENT APPLICATION NUMBER: US/09/890,688
;  CURRENT FILING DATE:  2001-08-06
;  PRIOR APPLICATION NUMBER: JP 11-346863
;  PRIOR FILING DATE: 1999-12-06
;  PRIOR APPLICATION NUMBER: JP 11-34684
;  PRIOR FILING DATE: 1999-12-06
;  PRIOR APPLICATION NUMBER: JP 2000-31062
;  PRIOR FILING DATE: 2000-02-08
;  PRIOR APPLICATION NUMBER: JP 2000-34091
;  PRIOR FILING DATE: 2000-02-10
;  PRIOR APPLICATION NUMBER: JP 2000-34090
;  PRIOR FILING DATE: 2000-02-10
;  PRIOR APPLICATION NUMBER: JP 2000-35829
;  PRIOR FILING DATE: 2000-02-14
;  PRIOR APPLICATION NUMBER: JP 2000-35899
;  PRIOR FILING DATE: 2000-02-14
;  PRIOR APPLICATION NUMBER: JP 2000-71161
;  PRIOR FILING DATE: 2000-03-14
;  PRIOR APPLICATION NUMBER: JP 2000-160851
;  PRIOR FILING DATE: 2000-05-30
;  NUMBER OF SEQ ID NOS: 160
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 82
;   LENGTH: 233
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-890-688-82

  Query Match             9.4%;  Score 113;  DB 3;  Length 233;
  Best Local Similarity   27.1%;  Pred. No. 0.0083;
  Matches   52;  Conservative   30;  Mismatches   88;  Indels   22;  Gaps    8;

Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                 |:|  |:     ::   |: || ::   :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
              |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209
```

```
Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221
```

RESULT 8
US-10-408-765A-969
; Sequence 969, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 969
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-969

Query Match 9.4%; Score 113; DB 4; Length 233;
Best Local Similarity 27.1%; Pred. No. 0.0083;
Matches 52; Conservative 30; Mismatches 88; Indels 22; Gaps 8;

```
Qy         31 ASFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADS 90
              ||  :||  |  |    : :  | |  || ||  :       ||  || |: |   | ||
Db         49 ASMWSSMSEDASV----ADMERAQL--EEEAAAAE-------ERPLVFLCSGCRRPLGDS 95

Qy         91 VHLAWDLSR-SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLY 149
                |:|  |:     ::   |: || ::    :       |     | |  | : :|:
Db         96 --LSWVASQEDTNCILLRCVSCNVSVDKEQKLSKREKENGCVLETLCCAGCSLNLGYVYR 153

Qy        150 STHAALAALRGHFCLSSDKMVCYLL--KTKAIVNASEMDIQNVPLSEKIAELKEKIVLTH 207
               |   |   |  |||| : :  |:|    | || : : :: |:   |   |::: :
Db        154 CTPKNLDYKRDLFCLSVEAIESYVLGSSEKQIV-SEDKELFNL---ESRVEIEKSLTQME 209

Qy        208 NRLKSLMKILSE 219
              : ||:|   | |
Db        210 DVLKALQMKLWE 221
```

RESULT 9
US-10-437-963-195427
; Sequence 195427, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 195427
; LENGTH: 708
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_91377C.1.pep
US-10-437-963-195427

Query Match 8.0%; Score 95.5; DB 4; Length 708;
Best Local Similarity 24.5%; Pred. No. 2.3;
Matches 71; Conservative 32; Mismatches 86; Indels 101; Gaps 17;

```
Qy          6 LRHRSRCATPPRGDF-------------------------CGGTERAIDQASFTTSMEW 39
              |||| : |||   |                           |||  |       |::
Db        225 LRHRRKAATPSASDASPPPPRRQSIVTIGEKEARAKAARAKSGGTSSA-----SPTTVST 279

Qy         40 DTQVVKGSSPLGPAGLGAEEPAAGPQLP----SW--LQPERCAVF---------QCAQCH 84
              |  | ||  : |:|    :|| || ||    :|  || |  :           : ::
Db        280 DVVPVVGSQEVTPSG-PISDPAGGPSLPEAVLTWEELQVEMGRLLKAGARGIGREISEAR 338

Qy         85 AVLADSVHLAWDLSRSLG----------AVVFSRVTNNVVLE------APFLVGIEGSLK 128
              |  | :  ||  | | |           :|      :  ||          |  | |||:
Db        339 AETAAANALAERLVRELAEAREDLTKMRELVAGNERQSKGLEDRMSELGDNLSEIRGSLR 398

Qy        129 GSTYNLL--FCGSCG----IPVG---FHLYSTHAALAALRG-----HFCLSSDKM----- 169
                 ||  |      ||    |||    | | |: | ||   |     |    :::
Db        399 -VTYTGLHQLAKECGVKSTIPVNLDEFSLTSSLAELATAMGEIPSKHTSRIAEETSNGIY 457

Qy        170 --VCYLLKTKAIVNAS--EMDIQNV----------PLSEKIAELKEKIV 204
               ||::|   | |  |  |:|:: :          : |::::| | ::
Db        458 TGVCHVL---ACVRLSRPELDLREILDQGAASDTRKEVMEEVSDLGESVL 504
```

RESULT 10
US-10-437-963-154548
; Sequence 154548, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 154548
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_54398C.1.pep
US-10-437-963-154548

Query Match 7.7%; Score 92; DB 4; Length 361;
Best Local Similarity 25.3%; Pred. No. 2.1;
Matches 61; Conservative 28; Mismatches 92; Indels 60; Gaps 14;

```
Qy          2 AAQPLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEPA 61
              | || |          | ||| |    : |     | || ||    |:|  | :|
Db         49 APTPLPFRRAVRAKAAQDGSGGTSSASPAVAST-----DVVVVPGSREATPSG-PASDPV 102

Qy         62 AGPQLP----SW--LQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVFSRVTNNVVL 115
              ||   |    ||  || |   : :      |:   :  |  |   :    | : ||
Db        103 AGRGSPAAVLSWEELQVEMGRLLEAGA--RVIGREIAEARGLEHRM-----SELGNN--- 152

Qy        116 EAPFLVGIEGSLKGSTYNLL--FCGSCGI-------PVGFHLYSTHAALAAL-----RGH 161
                 | |||:  ||  |    | |||        |  | | |: | |||        |
Db        153 ----LSEIRGSLR-VTYTGLHQLAGKCGIKSTIPANPDEFSLTSSLAELAAAMEEIPSKH 207
```

```
Qy        162 PCLSSDKM-------VCYLLKTKAIVNASEMDIQNV-----------PLSEKIAELKEKI 203
                  ::|       |::|    : :  |:|:: :           : |:: :| : :
Db        208 AARIGEEMSNRIYIGACHILACVRLAH-PELDLREILDQGEASDARKDVMEEVGDLGKSV 266

Qy        204 V 204
              :
Db        267 L 267
```

RESULT 11
US-10-369-493-18389
; Sequence 18389, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18389
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Lactococcus lactis
US-10-369-493-18389

Query Match 7.7%; Score 92; DB 4; Length 695;
Best Local Similarity 25.3%; Pred. No. 5.2;
Matches 38; Conservative 28; Mismatches 50; Indels 34; Gaps 7;

```
Qy        101 LGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRG 160
                 || ::|           ||  | :| ||     ::   : || | :  || :| : :| |
Db        101 LGTIIF------FYSGTPFFSGAKGELKSRKPAMMMLITMGITVAY-AYSVYATIMSLNG 153

Qy        161 HF---------CLSSDKMVCYLLKTKAIVNASE--MDI-------------QNVPLSE-K 195
                 |           |    :: :|:: |||: | :   |:              ::| ||| |
Db        154 HMGMNFWFELATLIVIMLIGHLIEMKAIMGAGDALKDLASLVPKKAHLKSGKDVELSELK 213

Qy        196 IAELKEKIVLTHNRLKSLMKILSEVTPDQS 225
                : :|   :|  : :: :   ||||   |:|
Db        214 VGDL--LLVKENEKIPADGLILSEALVDES 241
```

RESULT 12
US-09-780-525-2
; Sequence 2, Application US/09780525
; Patent No. US20020004223A1
; GENERAL INFORMATION:
; APPLICANT: Bin-Bing Zhou
; APPLICANT: Yuan Zhu
; APPLICANT: Priya Chaturvedi
; APPLICANT: Mark R. Hurle
; APPLICANT: Xiaotong Li
; TITLE OF INVENTION: FHAR1, A NEW RING FINGER PROTEIN
; FILE REFERENCE: GP-70668-C1
; CURRENT APPLICATION NUMBER: US/09/780,525
; CURRENT FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 09/456,876
; PRIOR FILING DATE: 1999-12-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 664
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-780-525-2

Query Match 7.4%; Score 89; DB 3; Length 664;
Best Local Similarity 23.0%; Pred. No. 9.9;
Matches 52; Conservative 26; Mismatches 74; Indels 74; Gaps 11;

```
Qy          8 HRSRCATPPRGDFC------GGTERAIDQA-----SFTTSMEWDTQVVKGSSPL------ 50
                :| : | ||    |       |  :|:  |     | ||:::    ::||  |
Db        434 YRRQAAQPPH---CPAPEGEPGAPQALGDAPSTSVSLTTAVQDYVCPLQGSHALCTCCFQ 490

Qy         51 ---GPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSR-------- 99
                    ::|   ||              |||  |   |   || | :|
Db        491 PMPDRRAEREQDPRVAPQ------------QCAVC---LQPFCHLYWGCTRTGCYGCLA 534

Qy        100 -----SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTY-NLLFCGSCGIPVGFHLYSTHA 153
                  :||      | ||   |:  |     : :| |: |:|               | :
Db        535 PFCELNLGDKCLDGVLNNNSYESDILKNYLAT-RGLTWKNML--------------TES 578

Qy        154 ALAALRGHFCLS-----SDKMVCYLLKTKAIVNASEMDIQNVPLSE 194
               :|  || | ||      | ::||    ::    :    ||:| ||
Db        579 LVALQRGVFLLSDYRVTGDTVLCYCCGLRSFRELTYQYRQNIPASE 624
```

RESULT 13
US-10-437-963-186569
; Sequence 186569, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 186569
; LENGTH: 435
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_83357C.1.pep
US-10-437-963-186569

Query Match 7.4%; Score 88.5; DB 4; Length 435;
Best Local Similarity 22.9%; Pred. No. 6.2;
Matches 50; Conservative 27; Mismatches 82; Indels 59; Gaps 12;

```
Qy          2 AAQPL--RH--RSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGA 57
                | :||  ||    |||         ||  |  :: |            :  || |  |
Db         41 APEPLSCRHGRHLRCAAVD-----GGAGRETERPSPPAPQ-------REESPSGSLGAAL 88

Qy         58 EEPAAGPQLPSWLQPERCAVFQCAQCHA--VLADSVHLAWDLSRSLGAVVFSRVTNNVVL 115
               |:|:  | :|  | |       | :|:|  : :: |    ||    :       : :| :
Db         89 EDPSPQPGVPLILLP------LCCRCYAKEICSEYVVRTTDLVNHI-------LNSNAIS 135

Qy        116 EAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLK 175
               | ||          |     | ||      | :  |           | ::|   || |:
Db        136 EGPF----------SMRKARFLGSAS---AFSVKQTEWP--------CATTDSKKCY-LQ 173

Qy        176 TKAIVNASEMDIQN------VPLSEKIAELKEKIVLTH 207
                 :    :| : |:       |  | : : ||  :  |
Db        174 NGSFGGITEDEEQSSLYNFLYPSKELLPDDKEMSIFDH 211
```

RESULT 14
US-10-437-963-155606

```
; Sequence 155606, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
;  APPLICANT: La Rosa, Thomas J.
;  APPLICANT:  Kovalic, David K.
;  APPLICANT:  Zhou, Yihua
;  APPLICANT:  Cao, Yongwei
;  APPLICANT:  Wu, Wei
;  APPLICANT:  Boukharov, Andrey A.
;  APPLICANT:  Barbazuk, Brad
;  APPLICANT:  Li, Ping
;  TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION:  Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53221)B
;  CURRENT APPLICATION NUMBER: US/10/437,963
;  CURRENT FILING DATE:  2003-05-14
;  NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 155606
;   LENGTH: 848
;   TYPE: PRT
;   ORGANISM: Oryza sativa
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(848)
;   OTHER INFORMATION: unsure at all Xaa locations
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT4530_55353C.1.pep
US-10-437-963-155606

  Query Match             7.3%;  Score 87.5;  DB 4;  Length 848;
  Best Local Similarity   22.1%;  Pred. No. 20;
  Matches   65;  Conservative   23;  Mismatches  109;  Indels   97;  Gaps   13;

Qy         16 PRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLGAEEP---------------- 60
              |||    || : : | :        |  :|||  ||| | |   |
Db         85 PRGFRLDGTGQGVRQTASCHLTRDLTDFLKGSGCLGPRGRGKRNPTPPPLGGNQEEGRHL 144

Qy         61 AAGPQLPSWLQPERCAVFQCAQCH-----------------AVLADSVHLAWDLSR---- 99
                || :|   :|        | |                     :  | |    |:|
Db        145 TLGPDVPEGARPRGRLCLPQASGHDTPGPISPTSLVLNKRTRIEIDEVVATMDISEANEG 204

Qy        100 --SLGAVV-FSRVTNNV-----VLEAPFLVGIEGSLK-GSTYNLLFCG------------ 138
               | |:|:  ||           |   ||    | ||   |||
Db        205 YVSCGSVIEMSRQMKAARAGVRAQSXPACXGIPTVGKVGSV--LLFTARMGEGLCHMFTT 262

Qy        139 -----SCGIPV-----------------GFHLYSTH---AALAALRGHFCLSSDKMVCYL 173
                  ||: :                 |:     |      | | || |   | : |
Db        263 GSKAMECGVKMCLVGWPWLAWDELGPRSGYQFGLNHRRPLVKAVLDGHLCEEEDAVNCLP 322

Qy        174 LKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLKSLMKILSEVTPDQSKP 227
                | : :|||             |  :  |    :::| :|:|| |: | | |
Db        323 KLRKTVGSASE-----------AKEAVKPAVKQEKKIRS-VKVLSLVS-DLSLP 363


RESULT 15
US-10-282-122A-51802
; Sequence 51802, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT:  Zamudio, Carlos
;  APPLICANT:  Malone, Cheryl
;  APPLICANT:  Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari
;  APPLICANT:  Zyskind, Judith
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John
;  APPLICANT:  Carr, Grant
;  APPLICANT:  Yamamoto, Robert
;  APPLICANT:  Forsyth, R.
;  APPLICANT:  Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE:  2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 51802
;   LENGTH: 669
;   TYPE: PRT
;   ORGANISM: Clostridium acetobutylicum
US-10-282-122A-51802

  Query Match             7.2%;  Score 86.5;  DB 4;  Length 669;
  Best Local Similarity   27.4%;  Pred. No. 18;
  Matches   37;  Conservative   22;  Mismatches   45;  Indels   31;  Gaps    8;

Qy         96 DLSR---SLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNL------LFCGSCGIPVGF 146
              |: |    :|: || | :|:|:    | ::|:    :| |  :      :|||  :  |
Db        366 DIGRKKVKIGSRVFVRRSNDVI---PEIMGVTEETEGETNEIEAPTICPYCGSEIVKEGV 422

Qy        147 HLYSTHAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVP-LSEKIAE-LKEKIV 204
              ||              ||   : : |     |:||: :  :  |:   ||| || | ||
Db        423 HL-------------FC--ENTLSCKPQMVKSIVHFASREAMNIEGFSEKTAEQLFEK-- 465

Qy        205 LTHNRLKSLMKILSE 219
              |    :  | :|  |
Db        466 LNIKSISDLYRITKE 480


Search completed: May 18, 2006, 15:56:08
Job time : 183 secs
```

GenCore version 5.1.8

OM protein - protein search, using sw model

Run on: May 18, 2006, 15:53:12 ; Search time 10 Seconds
(without alignments)
48.917 Million cell updates/sec

Title: US-09-942-052A-728
Perfect score: 1198
Sequence: 1 MAAQPLRHRSRCATPPRGDF.....LKSLMKILSEVTPDQSKPEN 229

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 21570 seqs, 2136119 residues

Total number of hits satisfying chosen parameters: 21570

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA_New:*
1: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
2: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
3: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
4: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
5: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
6: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
7: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
8: /EMC_Celerra_SIDS3/ptodata/1/pubpaa/US60_NEW_PUB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	71.5	6.0	1006	6	US-10-511-937-2425	Sequence 2425, Ap
2	67	5.6	388	7	US-11-312-958-44	Sequence 44, Appl
3	67	5.6	672	6	US-10-370-959-152	Sequence 152, App
4	66.5	5.6	345	7	US-11-181-115-5	Sequence 5, Appli
5	66.5	5.6	345	7	US-11-181-115-8	Sequence 8, Appli
6	66.5	5.6	345	7	US-11-181-115-12	Sequence 12, Appl
7	66.5	5.6	1722	6	US-10-505-928-780	Sequence 780, App
8	66	5.5	836	6	US-10-509-131-35	Sequence 35, Appl
9	65	5.4	331	7	US-11-118-524-9	Sequence 9, Appli
10	65	5.4	969	6	US-10-505-928-94	Sequence 94, Appl
11	63.5	5.3	5738	6	US-10-505-928-150	Sequence 150, App
12	63	5.3	555	7	US-11-242-505A-45	Sequence 45, Appl
13	62	5.2	314	6	US-10-538-066-366	Sequence 366, App
14	61.5	5.1	673	7	US-11-101-316-16	Sequence 16, Appl
15	61.5	5.1	799	6	US-10-505-928-716	Sequence 716, App
16	61	5.1	489	6	US-10-196-749-116	Sequence 116, App
17	61	5.1	642	6	US-10-505-928-259	Sequence 259, App
18	60.5	5.1	240	6	US-10-511-937-2977	Sequence 2977, Ap
19	60.5	5.1	459	6	US-10-370-959-60	Sequence 60, Appl
20	60.5	5.1	575	6	US-10-511-937-2625	Sequence 2625, Ap
21	60.5	5.1	999	6	US-10-196-749-434	Sequence 434, App
22	60.5	5.1	1821	6	US-10-505-928-451	Sequence 451, App
23	60	5.0	374	6	US-10-505-928-46	Sequence 46, Appl
24	60	5.0	382	6	US-10-196-749-20	Sequence 20, Appl
25	60	5.0	492	6	US-10-505-928-642	Sequence 642, App
26	59.5	5.0	738	6	US-10-511-937-2418	Sequence 2418, Ap
27	59	4.9	484	6	US-10-196-749-280	Sequence 280, App
28	59	4.9	484	7	US-11-101-316-78	Sequence 78, Appl
29	59	4.9	972	7	US-11-255-147-8	Sequence 8, Appli
30	59	4.9	1170	6	US-10-511-937-3007	Sequence 3007, Ap
31	58.5	4.9	769	7	US-11-246-999-67	Sequence 67, Appl
32	57.5	4.8	780	6	US-10-505-928-102	Sequence 102, App
33	57	4.8	481	7	US-11-249-111-77	Sequence 77, Appl
34	57	4.8	706	6	US-10-511-937-2450	Sequence 2450, Ap
35	57	4.8	728	7	US-11-311-555-18	Sequence 18, Appl
36	57	4.8	728	7	US-11-311-561-18	Sequence 18, Appl
37	57	4.8	2202	6	US-10-488-015-12	Sequence 12, Appl
38	57	4.8	4834	6	US-10-505-928-827	Sequence 827, App
39	56.5	4.7	264	6	US-10-196-749-12	Sequence 12, Appl
40	56	4.7	542	7	US-11-261-673-2	Sequence 2, Appli
41	55.5	4.6	638	7	US-11-302-678-47	Sequence 47, Appl
42	55.5	4.6	988	7	US-11-312-958-18	Sequence 18, Appl
43	55.5	4.6	2026	6	US-10-505-928-831	Sequence 831, App
44	55	4.6	408	6	US-10-505-928-455	Sequence 455, App
45	55	4.6	432	6	US-10-196-749-330	Sequence 330, App

ALIGNMENTS

RESULT 1
US-10-511-937-2425
; Sequence 2425, Application US/10511937
; Publication No. US20060088836A1
; GENERAL INFORMATION:
; APPLICANT: EXPRESSION DIAGNOSTICS, INC.
; APPLICANT: Wohlgemuth, Jay
; APPLICANT: Fry, Kirk
; APPLICANT: Woodward, Robert
; APPLICANT: Ly, Ngoc
; APPLICANT: Prentice, James
; APPLICANT: Morris, MacDonald
; APPLICANT: Rosenberg, Steven
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND MONITORING TRANSPLANT REJECTION
; FILE REFERENCE: 506612000104
; CURRENT APPLICATION NUMBER: US/10/511,937
; CURRENT FILING DATE: 2004-10-19
; PRIOR APPLICATION NUMBER: PCT/US2003/012946
; PRIOR FILING DATE: 2003-04-24
; PRIOR APPLICATION NUMBER: US 10/131,831
; PRIOR FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: US 10/325,899
; PRIOR FILING DATE: 2002-12-20
; NUMBER OF SEQ ID NOS: 3117
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2425
; LENGTH: 1006
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-511-937-2425

Query Match 6.0%; Score 71.5; DB 6; Length 1006;
Best Local Similarity 21.1%; Pred. No. 4.4;
Matches 56; Conservative 30; Mismatches 79; Indels 101; Gaps 14;

```
Qy          4 QPLRHRSRCATPPRGDF----------CGGTERAIDQASFTTSMEWDTQVVKGSSPLGP 52
              || |    |   ||| :          |     | : |                :|: |
Db        291 QPARGDKACQACPRGLYKSSAGNAPCSPCPARSHAPNPA---------------APVCP 334

Qy         53 AGLG--------AEEPAAGPQLPS-----WLQPERCAVFQCAQCHAVLADSVHLAWDLSR 99
                |         | |  ||  ||      | :               |   :: | | | |
Db        335 CLEGFYRASSDPPEAPCTGP--PSAPQELWFE-------------VQGSALMLHWRLPR 378

Qy        100 SLGA---VVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGIPVGFH-----LYST 151
               ||    ::|     ||| :       ||  : ::     | |   | |      |  :
Db        379 ELGGRGDLLF-----NVVCKE-----CEGRQEPASGGGGTCHRCRDEVHFDPRQRGLTES 428
```

```
Qy        152 HAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLK 211
Db        429 RVLVGGLRAH--------VPYILEVQAVNGVSELS----PDPPQAAAIN---VSTSHEVP 473

Qy        212 SLMKILSEVT------------PDQS 225
Db        474 SAVPVVHQVSRASNSITVSWPQPDQT 499
```

RESULT 2
US-11-312-958-44
; Sequence 44, Application US/11312958
; Publication No. US20060100152A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Rosenfeld, Julie Beth
; APPLICANT: Silos-Santiago, Inmaculada
; TITLE OF INVENTION: METHODS AND COMPOSITIONS IN TREATING
; TITLE OF INVENTION: PAIN AND PAINFUL DISORDERS USING 9949, 14230, 760, 62553,
; TITLE OF INVENTION: 12216, 17719, 41897, 47174, 33408, 10002, 16209, 314, 636,
; TITLE OF INVENTION: 27410, 33260, 619, 15985, 69112, 2158, 224, 615, 44373,
; TITLE OF INVENTION: 95431, 22245, 2387, 16658, 55054, 16314, 1613, 1675, 9569 OR
; TITLE OF INVENTION: 13424 MOLECULES
; FILE REFERENCE: MPI02-027P1RNOMNIM
; CURRENT APPLICATION NUMBER: US/11/312,958
; CURRENT FILING DATE: 2005-12-20
; PRIOR APPLICATION NUMBER: US/10/369,022
; PRIOR FILING DATE: 2003-02-19
; PRIOR APPLICATION NUMBER: US 60/360,495
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/370,121
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/373,010
; PRIOR FILING DATE: 2002-04-16
; PRIOR APPLICATION NUMBER: US 60/373,908
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/377,717
; PRIOR FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US 60/379,949
; PRIOR FILING DATE: 2002-05-13
; PRIOR APPLICATION NUMBER: US 60/382,409
; PRIOR FILING DATE: 2002-05-21
; PRIOR APPLICATION NUMBER: US 60/385,280
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 60/386,879
; PRIOR FILING DATE: 2002-06-06
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 388
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-312-958-44

Query Match 5.6%; Score 67; DB 7; Length 388;
Best Local Similarity 25.0%; Pred. No. 3.7;
Matches 29; Conservative 15; Mismatches 42; Indels 30; Gaps 4;

```
Qy         49 PLGPAGLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSV---HLAWDLSRSLGAVV 105
Db         59 PLPPPGL----------TPERLHARRQLYAACAVCFVFMAGEVVGGYLAHSLA------- 101

Qy        106 FSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFC---GSCGIPVGFHLYSTHAALAAL 158
Db        102 -------IMTDAAHLLADVGSMMGSLFSLWLSTRPATRTMTFGWHRSETLGALASV 150
```

RESULT 3
US-10-370-959-152
; Sequence 152, Application US/10370959
; Publication No. US20060088907A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Williamson, Mark W.
; APPLICANT: Kapeller-Libermann, Rosana
; APPLICANT: MacBeth, Kyle J.
; APPLICANT: Hunter, John J.
; APPLICANT: Rudolph-Owen, Laura A.
; APPLICANT: Bandaru, Rajasekhar
; APPLICANT: Tsai, Fong-Ying
; TITLE OF INVENTION: NOVEL 13237, 18480, 2245, 16228, 7677,
; TITLE OF INVENTION: 26320, 46619, 33166, 16836, 46867, 21617, 55562, 39228,
; TITLE OF INVENTION: 62088, 46745, 23155, 21657, 42755, 32229, 22325, 46863 AND
; TITLE OF INVENTION: 32252 MOLECULES AND USES THEREFOR
; FILE REFERENCE: MPI03-017OMNIM
; CURRENT APPLICATION NUMBER: US/10/370,959
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: US 09/910,150
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/219,028
; PRIOR FILING DATE: 2000-07-18
; PRIOR APPLICATION NUMBER: US 10/251,507
; PRIOR FILING DATE: 2002-09-20
; PRIOR APPLICATION NUMBER: US 09/715,479
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 60/218,053
; PRIOR FILING DATE: 2000-07-13
; PRIOR APPLICATION NUMBER: US 09/644,929
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 60/212,439
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: US 09/892,870
; PRIOR FILING DATE: 2001-06-26
; PRIOR APPLICATION NUMBER: US 60/214,174
; PRIOR FILING DATE: 2000-06-26
; PRIOR APPLICATION NUMBER: US 09/775,117
; PRIOR FILING DATE: 2001-02-01
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 156
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 152
; LENGTH: 672
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-370-959-152

Query Match 5.6%; Score 67; DB 6; Length 672;
Best Local Similarity 20.6%; Pred. No. 8.1;
Matches 36; Conservative 31; Mismatches 60; Indels 48; Gaps 7;

```
Qy         95 WDLSRSLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYN--------------LLFCGSC 140
Db         66 WKFS----GIVFSRVYDEVVDTSKGIADVPEWFKGSRLNYAENLLRHKENDRVALYIARE 121

Qy        141 G----IPVGFHLYSTHAAL--AALRGHFCLSSDKMVCYLLKTK-------------AIVN 181
Db        122 GKEEIVKVTFEELRQEVALFAAAMRKMGVKKGDRVVGYLPNSEHAVEAMLAAASIGAIWS 181

Qy        182 ASEMDIQNVPLSEKIAELKEKIVLT----------HNRLKSLMKILSEVTPDQSK 226
Db        182 STSPDFGVNGVLDRFSQIQPKLIFSVEAVVYNGKEHNHMEKLQQVVKGL-PDLKK 235
```

RESULT 4
US-11-181-115-5
; Sequence 5, Application US/11181115
; Publication No. US20060088511A1
; GENERAL INFORMATION:
; APPLICANT: Dana Farber Cancer Center
; TITLE OF INVENTION: Cancer Therapy Sensitizer
; FILE REFERENCE: 7032/2072
; CURRENT APPLICATION NUMBER: US/11/181,115
; CURRENT FILING DATE: 2005-07-14
; PRIOR APPLICATION NUMBER: PCT/US04/000901

```
;  PRIOR FILING DATE: 2004-01-14
;  PRIOR APPLICATION NUMBER: US 60/440,009
;  PRIOR FILING DATE: 2003-01-14
;  NUMBER OF SEQ ID NOS: 81
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
;   LENGTH: 345
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-181-115-5

  Query Match             5.6%;  Score 66.5;  DB 7;  Length 345;
  Best Local Similarity   25.0%;  Pred. No. 3.6;
  Matches   16;  Conservative    8;  Mismatches   15;  Indels   25;  Gaps    2;

Qy         47 SSPLGPAGLGAEEPAAGP-------QLPSWLQPERCAV-----------------FQCA 81
              |:|: |  :|  || | |       :  :|:| : | |                 |:|
Db        278 SAPMAPPAVGGGEPCAAPCCCLPQEETVAWVQCDGCDVWFHVACVGCSIQAAREADFRCP 337

Qy         82 QCHA 85
              | |
Db        338 GCRA 341


RESULT 5
US-11-181-115-8
; Sequence 8, Application US/11181115
; Publication No. US20060088511A1
; GENERAL INFORMATION:
;  APPLICANT: Dana Farber Cancer Center
;  TITLE OF INVENTION: Cancer Therapy Sensitizer
;  FILE REFERENCE: 7032/2072
;  CURRENT APPLICATION NUMBER: US/11/181,115
;  CURRENT FILING DATE:  2005-07-14
;  PRIOR APPLICATION NUMBER: PCT/US04/000901
;  PRIOR FILING DATE: 2004-01-14
;  PRIOR APPLICATION NUMBER: US 60/440,009
;  PRIOR FILING DATE: 2003-01-14
;  NUMBER OF SEQ ID NOS: 81
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
;   LENGTH: 345
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-181-115-8

  Query Match             5.6%;  Score 66.5;  DB 7;  Length 345;
  Best Local Similarity   25.0%;  Pred. No. 3.6;
  Matches   16;  Conservative    8;  Mismatches   15;  Indels   25;  Gaps    2;

Qy         47 SSPLGPAGLGAEEPAAGP-------QLPSWLQPERCAV-----------------FQCA 81
              |:|: |  :|  || | |       :  :|:| : | |                 |:|
Db        278 SAPMAPPAVGGGEPCAAPCCCLPQEETVAWVQCDGCDVWFHVACVGCSIQAAREADFRCP 337

Qy         82 QCHA 85
              | |
Db        338 GCRA 341


RESULT 6
US-11-181-115-12
; Sequence 12, Application US/11181115
; Publication No. US20060088511A1
; GENERAL INFORMATION:
;  APPLICANT: Dana Farber Cancer Center
;  TITLE OF INVENTION: Cancer Therapy Sensitizer
;  FILE REFERENCE: 7032/2072
;  CURRENT APPLICATION NUMBER: US/11/181,115
;  CURRENT FILING DATE:  2005-07-14
;  PRIOR APPLICATION NUMBER: PCT/US04/000901
;  PRIOR FILING DATE: 2004-01-14
;  PRIOR APPLICATION NUMBER: US 60/440,009
;  PRIOR FILING DATE: 2003-01-14
;  NUMBER OF SEQ ID NOS: 81
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
;   LENGTH: 345
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-11-181-115-12

  Query Match             5.6%;  Score 66.5;  DB 7;  Length 345;
  Best Local Similarity   25.0%;  Pred. No. 3.6;
  Matches   16;  Conservative    8;  Mismatches   15;  Indels   25;  Gaps    2;

Qy         47 SSPLGPAGLGAEEPAAGP-------QLPSWLQPERCAV-----------------FQCA 81
              |:|: |  :|  || | |       :  :|:| : | |                 |:|
Db        278 SAPMAPPAVGGGEPCAAPCCCLPQEETVAWVQCDGCDVWFHVACVGCSIQAAREADFRCP 337

Qy         82 QCHA 85
              | |
Db        338 GCRA 341


RESULT 7
US-10-505-928-780
; Sequence 780, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research et al.
;  TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
;  FILE REFERENCE: 28967/39178
;  CURRENT APPLICATION NUMBER: US/10/505,928
;  CURRENT FILING DATE:  2004-08-27
;  PRIOR APPLICATION NUMBER: US 60/363,019
;  PRIOR FILING DATE: 2002-03-07
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: PatentIn 3.2
; SEQ ID NO 780
;   LENGTH: 1722
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-505-928-780

  Query Match             5.6%;  Score 66.5;  DB 6;  Length 1722;
  Best Local Similarity   22.5%;  Pred. No. 35;
  Matches   18;  Conservative    6;  Mismatches   15;  Indels   41;  Gaps    2;

Qy         35 TSMEWDTQVVKGSSPLGPAGLGAEEPAAGPQLPSWLQPERC------------------- 75
              | :||  |: || :               |: | |  |:|
Db        784 TKLEWVCQIPKGRT--------------PKTPDWYNPDRAGIHGPPLIIEGSEYWFVAD 828

Qy         76 -------AVFQCAQCHAVLA 88
                    ||  ||  |: ||
Db        829 LHLNYEEAVLYCASNHSFLA 848


RESULT 8
US-10-509-131-35
; Sequence 35, Application US/10509131
; Publication No. US20060089493A1
; GENERAL INFORMATION:
;  APPLICANT: Biogen Idec Inc.
;  APPLICANT:  McLachlan, Karen
;  APPLICANT:  Gately, Dennis
;  TITLE OF INVENTION: NOVEL GENE TARGETS AND LIGANDS THAT BIND THERETO FOR TREATMENT AN
;  TITLE OF INVENTION:  DIAGNOSIS OF COLON CARCINOMAS
;  FILE REFERENCE: 037003-0302886
;  CURRENT APPLICATION NUMBER: US/10/509,131
;  CURRENT FILING DATE:  2004-09-28
;  PRIOR APPLICATION NUMBER: US 60/367,727
;  PRIOR FILING DATE: 2002-03-28
;  PRIOR APPLICATION NUMBER: US 60/381,328
;  PRIOR FILING DATE: 2002-05-20
;  PRIOR APPLICATION NUMBER: US 60/386,747
;  PRIOR FILING DATE: 2002-06-10
```

```
;  PRIOR APPLICATION NUMBER: US 60/427,564
;  PRIOR FILING DATE: 2002-11-20
;  NUMBER OF SEQ ID NOS: 67
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
;   LENGTH: 836
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-509-131-35

  Query Match             5.5%;  Score 66;  DB 6;  Length 836;
  Best Local Similarity   22.0%;  Pred. No. 14;
  Matches   42;  Conservative   14;  Mismatches   43;  Indels   92;  Gaps    8;

Qy          5 PLRHRSRCATPPRGDFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPA---------- 53
              |  :|||  :| |    ||:                    |||  |||
Db        496 PFIYRSR--SPCRASEAGGS--------------------GSSGRGPALCFEGSPPPEE 532

Qy         54 -------GLGAEEPAAGPQLPSWLQPERCAVFQCAQCHAVLADSVHLAWDLSRSLGAVVF 106
                  | |  ||  ||  ||  | |::                  ::: |  : :
Db        533 LPAVHSHGAGRGEPWPGPASPSGDQVSTCSL-----------------EMNYSSNSSLE 574

Qy        107 SRVTNNVVLEAPFLVGIEGS----------LKGSTYNLLFCGSCGIP------------ 143
               |  |:   |    ||:| |          ||  : |  |  |  |
Db        575 HRGPNSSTSE----VGLEASPGAAPDLRRTWKGG-HELPSCACCCEPQPSPAGPSAGAAG 629

Qy        144 -----VGFHLY 149
                  :| |||
Db        630 SSTLFLGPHLY 640


RESULT 9
US-11-118-524-9
; Sequence 9, Application US/11118524
; Publication No. US20060088847A1
; GENERAL INFORMATION:
;  APPLICANT: GU, WEI
;  TITLE OF INVENTION: ARF-BP1 AS MEDIATOR OF P53-DEPENDENT AND INDEPENDENT TUMOR
;  TITLE OF INVENTION:  SUPPRESSION AND USES THEREOF
;  FILE REFERENCE: 19240-497US2
;  CURRENT APPLICATION NUMBER: US/11/118,524
;  CURRENT FILING DATE:  2005-04-29
;  PRIOR APPLICATION NUMBER: 60/610,506
;  PRIOR FILING DATE: 2004-09-15
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.3
; SEQ ID NO 9
;   LENGTH: 331
;   TYPE: PRT
;   ORGANISM: homo sapiens
US-11-118-524-9

  Query Match             5.4%;  Score 65;  DB 7;  Length 331;
  Best Local Similarity   20.9%;  Pred. No. 5;
  Matches   28;  Conservative   29;  Mismatches   45;  Indels   32;  Gaps    6;

Qy        101 LGAVVFSRVTNNVVLEAPF-------LVGIEGSLK--GSTYNLLFCGSCGIPVGFHLYST 151
              :| |:   : || :|:  |       |:| :|: :  | ::          ||   || :
Db         71 IGIVLGLAIYNNCILDVHFPMVVYRKLMGKKGTFRDLGDSH----------PV---LYQS 117

Qy        152 HAALAALRGHFCLSSDKMVCYLLKTKAIVNASEMDIQNVPLSEKIAELKEKIVLTHNRLK 211
                  |    |:  :  | |: : :        |: |:   |:   : |  :|| :|:   |
Db        118 LKDLLEYEGN--VEDDMMITFQI--------SQTDLFGNPMMYDLKENGDKIPITNENRK 167

Qy        212 SLMKILSEVTPDQS 225
                : : |:   ::|
Db        168 EFVNLYSDYILNKS 181


RESULT 10
US-10-505-928-94
; Sequence 94, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research et al.
;  TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
;  FILE REFERENCE: 28967/39178
;  CURRENT APPLICATION NUMBER: US/10/505,928
;  CURRENT FILING DATE:  2004-08-27
;  PRIOR APPLICATION NUMBER: US 60/363,019
;  PRIOR FILING DATE: 2002-03-07
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: PatentIn 3.2
; SEQ ID NO 94
;   LENGTH: 969
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-505-928-94

  Query Match             5.4%;  Score 65;  DB 6;  Length 969;
  Best Local Similarity   19.7%;  Pred. No. 23;
  Matches   46;  Conservative   24;  Mismatches   72;  Indels   92;  Gaps   14;

Qy         11 RCATPPRG-DFCGGTERAIDQA-----SFTTSMEWDTQVVKGSSPLGPAGLGAEEPAAG- 63
              ||    :| : |  : ||  |         |     | :| |       | |||  |:|
Db        745 RCRRCHKGCETC--SSRAATQCLSCRRGFYHHQEMNTCVT-----LCPAGFYADESQKNC 797

Qy         64 ----PQLPSWL-QPERCAV-------------------------FQCAQCHAVLA---- 88
                   |      : :||:| |                       :| :||
Db        798 LKCHPSCKKCVDEPEKCTVCKEGFSLARGSCIPDCEPGTYFDSELIRCGECHHTCGTCVG 857

Qy         89 ----DSVHLA-------WDLSRSLGA------------VVFSRVTNNVVLEAPFLVGIEG 125
                  : :| |        |    : |                |  |   | :
Db        858 PGREECIHCAKNFHFHDWKCVPACGEGFYPEEMPGLPHKVCRRCDENCL---------- 906

Qy        126 SLKGSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKMVCYLLKTKAI 179
              |  ||: |   |  |     ||    |    : :  | | ::|:  | ::|:  :
Db        907 SCAGSSRN---CSRC--KTGF----TQLGTSCITNHTCSNADETFCEMVKSNRL 951


RESULT 11
US-10-505-928-150
; Sequence 150, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
;  APPLICANT: Ludwig Institute for Cancer Research et al.
;  TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
;  FILE REFERENCE: 28967/39178
;  CURRENT APPLICATION NUMBER: US/10/505,928
;  CURRENT FILING DATE:  2004-08-27
;  PRIOR APPLICATION NUMBER: US 60/363,019
;  PRIOR FILING DATE: 2002-03-07
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: PatentIn 3.2
; SEQ ID NO 150
;   LENGTH: 5738
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-505-928-150

  Query Match             5.3%;  Score 63.5;  DB 6;  Length 5738;
  Best Local Similarity   24.6%;  Pred. No. 4.2e+02;
  Matches   52;  Conservative   18;  Mismatches   68;  Indels   73;  Gaps   14;

Qy          8 HRSR---CATPP---RGDFCGGT--ERA---IDQASFTTS---MEWDTQVVKGSSPLGPA 53
              |:||   |  ||     |  | |   |||   :   | |    : |      | | :|
Db       3091 HQSRQRSCVDPPPKNGGAPCPGASQERAPCGLQPCSGGTGKGVLGWG----HGGSTVGTG 3146

Qy         54 GLGAEEPAAGPQLPSWLQPERCAVFQC-AQCHAV--------------LADSVHLAWDLS 98
                ||    ||  |:|     | |      |  :| :|                    | | |
Db       3147 RLGL--PA--PRLTWCPSPTRLRAGPCVCECRSVPEGAGAPMPTLLPGSQGQQKLQWALC 3202

Qy         99 RSLGAVVF-------------SRVTNNVVL---------EAPFLVGIEGSLKGSTYNL-- 134
                || :::               |    ::|          | | ||| |    | :: |
Db       3203 GSLPSLLCPLGLSALFHILPGCRCPPGLLLHDTRCLPLSECPCLVGEELKWPGVSFLLGN 3262
```

```
Qy        135 ----------LFC--GSCGIPVGFHLYSTHA 153
                        | |  | | :| |:  :|: |
Db       3263 CSQCVCEKGELLCQPGGCPLPCGWSAWSSWA 3293
```

RESULT 12
US-11-242-505A-45
; Sequence 45, Application US/11242505A
; Publication No. US20060099656A1
; GENERAL INFORMATION:
; APPLICANT: Carroll, Joseph M.
; APPLICANT: Healy, Aileen
; TITLE OF INVENTION: Methods and Compositions for Treating
; TITLE OF INVENTION: Hematological Disorders Using 232, 2059, 10630, 12848, 13875,
; TITLE OF INVENTION: 14395, 14618, 17692, 58874, 252, 304, 1980, 14717, 9941, 19310,
; FILE REFERENCE: MPI2001-288P1RCP1OMNIM
; CURRENT APPLICATION NUMBER: US/11/242,505A
; CURRENT FILING DATE: 2005-10-03
; PRIOR APPLICATION NUMBER: US 10/290,078
; PRIOR FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: US 60/347,949
; PRIOR FILING DATE: 2001-11-07
; PRIOR APPLICATION NUMBER: US 10/320,351
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: 60/341,606
; PRIOR FILING DATE: 2001-12-17
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 45
; LENGTH: 555
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-242-505A-45

Query Match 5.3%; Score 63; DB 7; Length 555;
Best Local Similarity 26.1%; Pred. No. 18;
Matches 18; Conservative 6; Mismatches 27; Indels 18; Gaps 3;

```
Qy         37 MEWDTQVVKGSSP------------LGPAGLGAEEPAAGPQLPSWLQPER-CAVFQCAQ 82
                 : ||    :   |              | | | |||  |     | : |:   | :|
Db        264 IHWDQASARPRGPEIEPRGEGDHNHDTGEGGQGGEEPRGG----RWDEDEQWSVVVECED 319

Qy         83 CHAVLADSV 91
              |  : || |
Db        320 CELIPADHV 328
```

RESULT 13
US-10-538-066-366
; Sequence 366, Application US/10538066
; Publication No. US20060094649A1
; GENERAL INFORMATION:
; APPLICANT: Epimmune Inc.
; TITLE OF INVENTION: HLA-A1, -A2, -A3, -A24, -B7, and -B44 Tumor Associated Antigen
; TITLE OF INVENTION: Peptides and Compositions
; FILE REFERENCE: 2060.015PC06
; CURRENT APPLICATION NUMBER: US/10/538,066
; CURRENT FILING DATE: 2005-06-09
; PRIOR APPLICATION NUMBER: US 60/432,017
; PRIOR FILING DATE: 2002-12-10
; NUMBER OF SEQ ID NOS: 767
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 366
; LENGTH: 314
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-538-066-366

Query Match 5.2%; Score 62; DB 6; Length 314;
Best Local Similarity 22.1%; Pred. No. 10;
Matches 40; Conservative 20; Mismatches 79; Indels 42; Gaps 7;

```
Qy          5 PLRHRSRCATPPRG--------DFCGGTERAIDQASFTTSMEWDTQVVKGSSPLGPAGLG 56
                || ||:   |  |           |    | ::    :|     :|  |  |   :
Db          2 PLEQRSQHCKPEEGLEARGEALGLVGAQAPATEEQEAASSSSTLVEVTLGEVPAAESPDP 61

Qy         57 AEEPAAGPQLPS------WLQ---------------PERCAVFQCAQCHAVLADSVHLA 94
                : |     ||:     | |               |:  : || |    | |: ||
Db         62 PQSPQGASSLPTTMNYPLWSQSYEDSSNQEEEGPSTFPDLESEFQAALSRKV-AELVHFL 120

Qy         95 WDLSRSLGAVVFSRVTNNVVLEAPFLVGIEGSLKGSTYNLLFCGSCGI------PVGFHL 148
                 |:   |  : :  :||    :   :  |   |:  |:|     ||      |:| ||
Db        121 LLKYRAREPVTKAEMLGSVVGNWQYFFPVIFSKASSSLQLVF----GIELMEVDPIG-HL 175

Qy        149 Y 149
              |
Db        176 Y 176
```

RESULT 14
US-11-101-316-16
; Sequence 16, Application US/11101316
; Publication No. US20060099657A1
; GENERAL INFORMATION:
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Grimaldi,Christopher J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Wood,William I.
; TITLE OF INVENTION: ANTIBODIES TO A POLYPEPTIDE ENCODED BY A NUCLEIC ACID
; TITLE OF INVENTION: UNDEREXPRESSED IN MELANOMA
; FILE REFERENCE: P3230R1C17C1
; CURRENT APPLICATION NUMBER: US/11/101,316
; CURRENT FILING DATE: 2005-04-06
; PRIOR APPLICATION NUMBER: 10/063526
; PRIOR FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: 10/006867
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: 09/380137
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: PCT/US99/12252
; PRIOR FILING DATE: 1999-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 16
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo Sapien
US-11-101-316-16

Query Match 5.1%; Score 61.5; DB 7; Length 673;
Best Local Similarity 28.7%; Pred. No. 34;
Matches 31; Conservative 7; Mismatches 43; Indels 27; Gaps 7;

```
Qy         10 SRCATPPRGDFCGGTERAIDQASF-------TTSMEWDTQVVKG----SSPLGPAGLGAE 58
                :||  ||:    |     :| | |       | ::     ||:     || | |  |
Db        326 TRCHFPPKN--AGRLLLELDYADFGCPATTTTATVPTTRPVVREPTALSSSLAPTWLSPT 383

Qy         59 EPAA-GPQLPSWL--------QPERCAVFQC---AQCHAVLADSVHLA 94
               ||   |  ||          ||: |    |     ||  |    |||
Db        384 APATEAPSPPSTAPPTVGPVPQPQDCPPSTCLNGGTCH--LGTRHHLA 429
```

RESULT 15
US-10-505-928-716
; Sequence 716, Application US/10505928
; Publication No. US20060088532A1
; GENERAL INFORMATION:
; APPLICANT: Ludwig Institute for Cancer Research et al.
; TITLE OF INVENTION: LYMPHATIC ENDOTHELIAL GENES
; FILE REFERENCE: 28967/39178
; CURRENT APPLICATION NUMBER: US/10/505,928
; CURRENT FILING DATE: 2004-08-27

```
;  PRIOR APPLICATION NUMBER: US 60/363,019
;  PRIOR FILING DATE: 2002-03-07
;  NUMBER OF SEQ ID NOS: 866
;  SOFTWARE: PatentIn 3.2
; SEQ ID NO 716
;   LENGTH: 799
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-505-928-716

  Query Match             5.1%;  Score 61.5;  DB 6;  Length 799;
  Best Local Similarity   29.0%;  Pred. No. 43;
  Matches   36;  Conservative    8;  Mismatches   45;  Indels   35;  Gaps    8;

Qy         65 QLPSWLQPERCAVFQCAQCH----------AVLADSVHLAWDL-SRSLGAVVFSRVT 110
              :|  | |||   :|  | |               : |:      | :||      || |
Db        383 ELSVWDQPEDLNLFFTATCQDGVSYPGQRKCEGLKIGDTASFEVSLEARSCP----SRHT 438

Qy        111 NNVVLEAPFLVGIEGSLK-GSTYNLLFCGSCGIPVGFHLYSTHAALAALRGHFCLSSDKM 169
               :|    |  ||   ||: | |||   | :||  ||    |              |  |
Db        439 EHVFALRP--VGFRDSLEVGVTYN---C-TCGCSVGLEPNSAR----------CNGSGTY 482

Qy        170 VCYL 173
              || |
Db        483 VCGL 486


Search completed: May 18, 2006, 15:56:23
Job time : 11 secs
```